

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER FAKULTÄT FÜR
CHEMIE UND PHARMAZIE DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Structural basis of human mitochondrial transcription initiation and processive elongation



Hauke Sven Hillen
aus
Erlangen, Deutschland
2017

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Göttingen, den 29.09.2017

Hauke Sven Hillen

Dissertation eingereicht am 09.10.2017

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 06.11.2017

Acknowledgements

First and foremost, I would like to thank Patrick Cramer for giving me the opportunity to work with him and pursue this challenging, yet rewarding project. His enthusiasm for structural biology immediately sparked over to me at our very first meeting. His immediate trust in me to manage this project, which involved an international collaboration, was highly motivating and rewarding to me. The collaborative atmosphere Patrick fosters in his lab is truly unique, and combined with the freedom granted by him to pursue my own ideas made my PhD research a deeply enjoyable time.

My second thanks go to Dmitry Temiakov, our collaborator on this project. None of the results presented here would have been possible without his input, ideas, critical judgement and, last but not least, his active participation in bench work. Thank you for always sharing your experience and knowledge (both practical and theoretical) on mitochondrial biology, for critically evaluating my experimental ideas and for putting up with my constant questions and inquiries. Lastly, I thank you for great times together at numerous scientific meetings and occasions inside and outside the “science world”. Thank you for becoming not just a collaborator, but a friend!

I am thankful to all the past and present members of the Cramer group, who have made my time in the lab such an enjoyable experience. Special thanks go to Kathrin Schwinghammer, who introduced me to the mitochondrial transcription project and got me started in the lab. Furthermore, I thank Simon Neyer for great conversations (science and non-science) and for being my Squash partner. I thank Carrie Bernecky for being a great “bay-neighbor” and, more importantly, a great person! (And for critically reading this thesis!) Furthermore, I thank Merle Hantsche, Carina Demel, Katharina Hofmann, Felix Wagner, Christian Dienemann, Tobi Gubbey and Sarah Sainsbury. I am indebted to Jürgen Wawrzinek for his service in the crystallization facility of the MPIbpc and to Claudia Buchen and Kerstin Maier for keeping our lab running. I further thank Sarah Sainsbury and Christoph Engel for their help with X-ray crystallography. I am also thankful to past and present members of the Temiakov lab.

I am very grateful to the Boehringer Ingelheim Fonds, who supported me with a PhD student fellowship. It was a great honor and the scientific as well as personal support was truly outstanding. Thank you Claudia Walther, Anja Hofmann and Sandra Schedler for your great administrative work and personal support!

I am thankful to my examination committee members for their support and interest in my work: Patrick Cramer, Dietmar Martin, Klaus Förstemann, Fabiana Perocchi, Karl-Klaus Conzelmann and Karl-Peter Hopfner. Thank you very much!

I thank my mother for her support, love and invaluable, experienced life advice.

Lastly, I thank Laura for her love, her understanding support and her honest and critical advice – thank you for being there for me!

Für Wolfgang.

Summary

In eukaryotic cells, mitochondria produce the vast majority of ATP, the universal energy currency of life. To do so, they maintain a highly reduced genome as well as the molecular machinery necessary for its expression. Transcription in mitochondria is carried out by a dedicated mitochondrial RNA polymerase (mtRNAP), which is related to single-subunit RNA polymerases (RNAPs) found in bacteriophages. In contrast to these self-sufficient enzymes, however, mtRNAP requires additional protein factors for all steps of transcription, suggesting a complex regulation. Moreover, it also produces the RNA primers necessary to initiate DNA synthesis, placing this enzyme at the heart of mitochondrial gene expression and organelle maintenance. Structures of mtRNAP have provided a first glimpse at the central actor orchestrating these important processes, but the mechanistic principles governing the individual steps of mitochondrial transcription remain poorly understood. In this study, we expand our understanding of these processes by investigating the structural basis of transcription initiation and processive elongation, two steps of regulatory importance.

To initiate transcription, mtRNAP associates with the two initiation factors TFAM and TFB2M and promoter DNA to form an initiation complex (IC). Here, I present the structure of human TFB2M at 1.75 Å resolution and of the human initiation complex at 4.5 Å resolution. Together with published structures of mtRNAP and TFAM, this allows for construction of a pseudo-atomic model of the IC. The structures reveal how mtRNAP is recruited to the promoter by TFAM and suggest that TFB2M induces a rearrangement in mtRNAP to facilitate promoter opening. The open complex is further stabilized by interactions between TFB2M and the melted non-template DNA strand. Structural comparisons demonstrate that transition to elongation is accompanied by a profound re-arrangement of the upstream DNA.

Following initiation, mtRNAP associates with the elongation factor TEFM for processive transcription elongation. This factor enables mtRNAP to transcribe through a G-quadruplex forming sequence in the mitochondrial genome, which otherwise leads to transcription termination and primer formation for replication. However, the mechanistic basis for this anti-termination activity of TEFM is unknown. Here, I present crystal structures of the human TEFM domains and, in a collaborative effort with the Temiakov Lab, we functionally define their roles in transcription. In addition, I have determined the structure of an anti-termination complex, comprised of the functional domains of TEFM bound to transcribing mtRNAP. These structures demonstrate that TEFM stabilizes the elongation complex by enclosing the downstream DNA in a “sliding clamp” and by interacting with the non-template strand in the transcription bubble. Moreover, these data suggest that TEFM prevents formation of the G-quadruplex in the RNA exit path, thereby mediating the switch between transcription and DNA replication.

Taken together, these results greatly advance our understanding of mitochondrial transcription and elucidate the mechanistic basis for the factor dependence of mtRNAP. Furthermore, they provide a framework for future studies aimed at deciphering the regulatory mechanisms of transcription and DNA replication in human mitochondria.

Publications

Parts of this work have been published or are currently in the process of publication:

H.S. Hillen, Y.I. Morozov, A. Sarfallah, D. Temiakov and P. Cramer (2017) Structural basis of mitochondrial transcription initiation. *Cell*, under review

Author Contributions: H.S.H. cloned constructs for structure determination, purified proteins, assembled the IC complexes, crystallized TFB2M and the IC, collected and analyzed diffraction data, solved the crystal structures and built and refined atomic models. Y.I.M. and A.S. cloned constructs and performed mutagenesis and functional assays. P.C. and D.T. designed and supervised research. H.S.H., D.T. and P.C. interpreted the data and wrote the manuscript.

H.S. Hillen, A.V. Parshin, K. Agaronyan, Y.I. Morozov, J.J. Graber, A. Chernev, K. Schwinghammer, H. Urlaub, M. Anikin, P. Cramer and D. Temiakov (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell*, in press

Author contributions: H.S.H. cloned and purified proteins, prepared crystals, solved structures and performed BS3 and EDC cross-linking and analytical size-exclusion chromatography. A.V.P. and J.J.G. made and purified TEFM mutants and performed Bpa cross-linking and mapping. K.A. and Y.I.M. isolated TEFM mutants, performed functional assays, and made and analyzed DSG cross-links. K.S. performed cloning and established initial purification of TEFM. A.C. and H.U. performed mass-spectrometric analysis of cross-linked complexes. D.T. purified proteins, performed foot-printing, and prepared TEFM NTD crystals. H.S.H., A.V.P., K.A., M.A., Y.I.M., P.C. and D.T. analyzed the data, and H.S.H., M.A., P.C. and D.T. wrote the manuscript.

Additional publication:

H.S. Hillen^{*}, D.A. Markov^{*}, I.D. Wojtas, K. Hofmann, A.T. Cowan, J.L. Jones, D. Temiakov, P. Cramer and M. Anikin (2017) The PPR protein-based mechanism of RNA maturation as an alternative to polyadenylation. (manuscript in preparation)

^{*} These authors contributed equally to this work.

Contents

Erklärung.....	III
Eidesstattliche Versicherung	III
Acknowledgements.....	IV
Summary	V
Publications	VI
Contents	VII
1. Introduction.....	1
1.1. The flow of genetic information	1
1.2. Transcription.....	1
1.3. Single-subunit RNA polymerases	2
1.4. Multi-subunit RNA polymerases	3
1.5. Mitochondria.....	4
1.6. Transcription in human mitochondria	6
1.6.1. Mitochondrial transcription initiation	6
1.6.2. Transcription and DNA replication are coupled in mitochondria	8
1.6.3. Mitochondrial transcription elongation.....	9
1.7. Previous structural studies of mitochondrial transcription.....	9
1.8. Aim and scope	13
2. Materials and Methods	14
2.1. Materials.....	14
2.1.1. Bacterial strains.....	14
2.1.2. Growth media for <i>E.coli</i> and additives	14
2.1.3. Plasmids	15
2.1.4. Primers	17
2.1.5. Synthetic oligonucleotides	19
2.1.6. Buffers and solutions.....	20
2.2. Methods.....	22
2.2.1. General methods	22
Site-directed mutagenesis PCR.....	22
Around-the-Horn PCR.....	23
Transformation of <i>E.coli</i> cells	23
Isolation of plasmid DNA and sequence verification.....	23
Preparation of glycerol stocks for protein expression	23
Protein and nucleic acid concentration determination.....	24
Protein identification by mass spectrometry.....	24
Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).....	24
2.2.2. Protein purification	24
Purification of human mitochondrial RNA polymerase	24
Purification of human TFAM.....	25

Purification of human TFB2M.....	26
Purification of human TEFM.....	27
Purification of selenomethionine-labeled proteins.....	27
2.2.3. General crystallization methods.....	28
Microseeding of crystals.....	28
Cryo-protection and freezing of crystals.....	28
2.3. Structural basis of mitochondrial transcription initiation.....	29
Protein expression constructs.....	29
Crystallization of human TFB2M.....	29
Reconstitution and crystallization of the IC.....	30
Data collection, structure determination and refinement.....	30
2.4. Mechanism of transcription anti-termination in human mitochondria.....	33
Expression and purification of the components of human mitochondrial transcription.....	33
Crystallization of human TEFM C-terminal domain.....	33
Crystallization of the EC/TEFM complex.....	34
Structure determination, model building and refinement.....	34
Mapping of TEFM-mtRNAP interaction using BS3 and EDC cross-linking.....	35
Analytical size-exclusion chromatography.....	36
3. Results and Discussion.....	37
3.1. Structural basis of mitochondrial transcription initiation.....	37
3.1.1. Abstract.....	37
3.1.2. Results.....	38
Structure of human TFB2M.....	38
Structure determination of the mitochondrial transcription initiation complex.....	41
IC structure reveals locations of TFAM, TFB2M and DNA on mtRNAP.....	43
TFAM recruits mtRNAP to promoter DNA.....	45
TFB2M assists mtRNAP in DNA opening.....	47
Comparison to T7 RNAP initiation.....	50
Transition from initiation to elongation.....	52
3.1.3. Discussion.....	53
3.2. Mechanism of transcription anti-termination in human mitochondria.....	55
3.2.1. Abstract.....	55
3.2.2. Results.....	56
TEFM contains two structured domains.....	56
Structure of TEFM.....	58
Functional role of TEFM domains.....	60
Structure determination of TEFM-mtRNAP elongation complex.....	62
Architecture of the TEFM-mtRNAP elongation complex.....	63
Implication for mtRNAP translocation.....	67
Mapping of TEFM-mtRNAP interactions by cross-linking.....	69
Mechanisms of transcription termination and anti-termination at CSBII.....	72
3.2.3. Discussion.....	74
The bipartite role of TEFM in transcription elongation and anti-termination.....	74
TEFM as a dedicated transcription factor.....	75
Target specificity of TEFM.....	76

4. Conclusion and Outlook	78
4.1. Towards high-resolution structural data on the mitochondrial IC	79
4.2. A complete picture of mitochondrial transcription initiation	80
4.3. Further studies on TEFM function and regulation.....	81
4.4. Evolutionary aspects of mitochondrial transcription.....	82
4.5. Clinical relevance of mtRNAP structures.....	82
5. Appendix	84
5.1. Extended Materials and Methods for the initiation study.....	84
5.1.1. Extended Methods for the initiation study	84
Transcription assays	84
5.1.2. Extended figures and tables of the initiation study	85
5.2. Extended Materials and Methods for the elongation study	89
5.2.1. Extended Methods for the elongation study	89
Limited proteolysis of TEFM and analysis of the cleavage products.....	89
Crystallization of human TEFM N-terminal domain.....	89
Promoter templates for transcription assays	89
Transcription assays	90
Nuclease foot-printing assays.....	90
Cross-linking using artificial photo reactive amino acid (Bpa)	90
Protein-DNA photo cross-linking	90
Mapping of TEFM-mtRNAP interaction using disuccinimidyl glutarate (DSG) cross-linking ..	91
Mapping of the Bpa-TEFM cross-linking sites in mtRNAP	91
Pyrophosphorolytic assay	91
5.2.2. Extended figures and tables of the elongation study	92
References	98
List of Abbreviations	109
List of Figures	112
List of Tables	113

1. Introduction

1.1. The flow of genetic information

The genetic information of all life on earth is encoded in nucleic acid. With few exceptions, the molecule in which information is stored is deoxyribonucleic acid (DNA). First isolated by Friedrich Miescher in 1869, it was not until the mid-20th century that DNA was identified as the carrier of hereditary information in famous experiments by Avery, MacLeod and McCarty (Avery et al., 1944) and later by Hershey and Chase (Hershey and Chase, 1952). Shortly after, the structure of DNA was unraveled by Watson and Crick, with the help of X-Ray fiber diffraction data produced by Wilkins, Stokes and Franklin (Watson and Crick, 1953; Wilkins et al., 2003). The emerging field of molecular biology was moving at a fast pace, and in 1958 led Francis Crick to postulate the so-called “central dogma of molecular biology”, which proposes a unidirectional flow of genetic information from nucleic acids to proteins (Crick, 1958). Specifically, Crick proposed that sequence information is passed from DNA on to RNA and from there to proteins (Crick, 1970) (Figure 1). Although some additional information transfers have been discovered since (RNA to DNA and RNA to RNA) (Baltimore, 1970; Baltimore and Franklin, 1962), this general concept still holds true today. The process of information transfer from DNA to RNA is called “transcription” and the information transfer from RNA to protein is referred to as “translation”. In multi-cellular organisms, all cells contain the same DNA. The specific fate and function of a cell, its phenotype, is thus determined by which genetic information is activated or repressed in the particular cell. Therefore, understanding the fundamental principles and mechanisms governing the expression of genetic information is prerequisite to understanding complex biology.

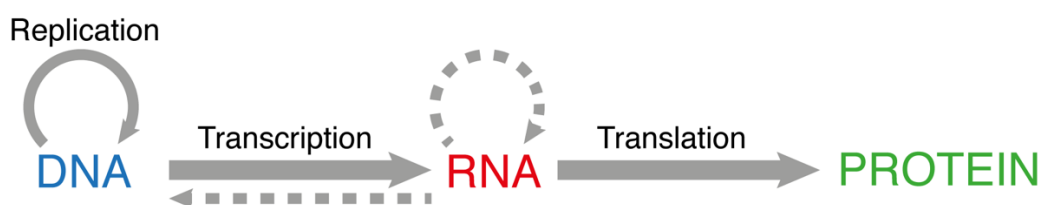


Figure 1. The central dogma of molecular biology.

The flow of genetic information is represented as arrows. Thick, solid arrows represent the main avenues of information flow during gene expression. The dashed arrows represent other directions of information flow that have been observed in nature (Baltimore, 1970; Baltimore and Franklin, 1962). Adapted from (Crick, 1970).

1.2. Transcription

It comes as no surprise that in all organisms, from unicellular life to mammals, gene expression is heavily regulated at the level of transcription, the first step in gene expression. Regulating expression of genetic information at this step is highly efficient in terms of resource usage and

nevertheless allows the cell to respond to changing cellular needs or environmental conditions. Chemically, the transfer of information encoded in DNA to RNA requires the synthesis of a linear polymer of ribonucleotides using the linear, polymeric deoxyribonucleotides of DNA as template. Such catalytic activity for template-dependent nucleotide polymerization is required for both replication and expression of genetic information, and thus it may be one of the most ancient enzymatic activities (Steitz, 1998). The enzymes which catalyze transcription are called DNA-dependent RNA polymerases, or short RNAPs, and can be found in all domains of life. However, the molecular architecture of the transcription apparatus differs significantly between organisms. RNAPs can be divided into two evolutionarily unrelated classes: single-subunit and multi-subunit RNAPs (Cramer, 2002a). However, structural comparison demonstrates that they share similarities in the topology of nucleic acid binding, suggesting that their mode of action has evolved through convergent evolution dictated by mechanistic principles (Cramer, 2002a).

The process of transcription can be divided into three distinct steps common to all types of RNA polymerases: initiation of RNA synthesis, chain elongation, and termination of transcription. During initiation, the RNA polymerase associates with the DNA template at a specific region within the genome called a promoter. Before RNA synthesis can commence, the two DNA strands must be unwound in order to expose the template strand around the transcription start site. Following initial ribonucleotide polymerization, the polymerase transitions to a productive elongation state, which is characterized by high fidelity and processivity. Chain elongation typically continues until a termination signal is encountered, which leads to dissociation of the RNA polymerase from the DNA template and the newly synthesized RNA (Hippel et al., 1984).

The various RNA polymerases found in nature utilize different strategies to achieve the aforementioned steps, ranging from small enzymes which act on a limited number of promoters and have all functions combined in one polypeptide to complex multi-subunit enzymes which rely on auxiliary factors for each step.

1.3. Single-subunit RNA polymerases

The simplest type of RNA polymerases are single-subunit DNA-dependent RNAPs (ssRNAPs), which consist of only one polypeptide chain and were initially identified in bacteriophages (Chamberlin et al., 1970). Structurally, these enzymes show similarity to the pol I family of DNA polymerases, which includes the *E.coli* DNA polymerase Klenow fragment and the Human Immunodeficiency Virus 1 reverse transcriptase (Kohlstaedt et al., 1992; McAllister and Raskin, 1993; Sousa, 1996). These proteins display a conserved fold of the catalytic domain resembling a right hand, consisting of palm, fingers and thumb subdomains (Figure 2) (Cheetham et al., 1999; Ollis et al., 1985; Sousa et al., 1993). Single-subunit RNAPs are widespread in nature, as sequences encoding ssRNAPs can be found not only in viruses but also in eukaryotic genomes and in linear plasmids in the mitochondria of certain fungi and plants (Cermakian et al., 1996; Kempken et al., 1992; Masters et al., 1987; Tiranti et al., 1997). Based on phylogenetic analysis, single-subunit RNAPs can be further sub-divided into three families: bacteriophage-encoded, eukaryotic nucleus-encoded and mitochondrial linear

plasmid-encoded (Cermakian et al., 1997). These three classes generally share a high degree of conservation in their C-terminal part, which constitutes the catalytic domain, while the N-terminal regions are highly divergent across species and families (Cermakian et al., 1997).

Arguably the most well-studied single-subunit RNAP is the bacteriophage T7 RNA polymerase (T7 RNAP), which has thus become a paradigm for single-subunit RNAP function. This roughly 100 kDa enzyme is capable of performing the entire transcription cycle without the requirement of any further protein factors. The mechanistic basis of transcription initiation and elongation by T7 RNAP has been elucidated by visualizing structural snapshots of each of the steps involved. To initiate transcription, T7 RNAP interacts with the promoter DNA upstream of the transcription start site via the “specificity loop”, a β -hairpin structure within its C-terminal domain, and the “AT-rich recognition loop” located in the N-terminal promoter binding domain (Cheetham and Steitz, 1999; Cheetham et al., 1999; Durniak et al., 2008). DNA strand separation is facilitated by the “intercalating hairpin”, which also guides the template strand towards the active site (Cheetham and Steitz, 1999; Cheetham et al., 1999). As initial RNA synthesis proceeds, the growing RNA-DNA hybrid induces a profound structural rearrangement of the N-terminal domain of T7 RNAP, effectively destroying the promoter binding domain and allowing the polymerase to transition to elongation (Durniak et al., 2008; Yin and Steitz, 2002). During elongation, the polymerase processively undergoes cycles of substrate binding, nucleotide addition and translocation of the RNA-DNA hybrid (Tahirov et al., 2002; Temiakov et al., 2004; Yin and Steitz, 2002; 2004).

1.4. Multi-subunit RNA polymerases

Transcription in prokaryotes, archaea and in the nuclei of eukaryotic cells is carried out by the more complex multi-subunit RNAPs, which, depending on the organism, are comprised of 5 – 15 subunits (Cramer, 2002b; Werner and Grohmann, 2011). The prokaryotic RNAP consists of a highly conserved five-subunit core with a total mass of approximately 400 kDa (Lane and Darst, 2010a). In order to initiate transcription, this core associates with a promoter-specific σ -factor to form the so-called “holoenzyme” (Saecker et al., 2011). The level of complexity is markedly increased in eukaryotic cells, in which three multi-subunit RNA polymerases can be found: RNA Polymerase I, II and III (Pol I – III) (Roeder and Rutter, 1969). While Pol I transcribes the highly structured ribosomal RNA precursors (pre-rRNA), Pol II is the enzyme which transcribes protein coding genes to produce messenger RNA (mRNA), and Pol III transcribes small and transfer RNAs (tRNAs) (Cramer, 2002b). In plants, two further multi-subunit RNAPs were found, Pol IV and Pol V, which are related to Pol II and are involved in the transcription of small interfering RNAs (siRNAs) (Haag and Pikaard, 2011; Ream et al., 2009).

Although composed of a much larger number of subunits, sequence analysis suggests that eukaryotic multi-subunit RNAPs are evolutionarily related to prokaryotic RNAPs and these enzymes share a common core (Lane and Darst, 2010b; Sweetser et al., 1987; Werner and Grohmann, 2011). This was confirmed around the turn of the millennium, when crystal structures of both bacterial RNAP as well as yeast RNA Pol II became available (Cramer et al.,

2001; 2000; Gnatt et al., 2001; Lane and Darst, 2010b; Zhang et al., 1999). The conserved core of multi-subunit RNAPs resembles a “crab claw” with a large cleft that accommodates nucleic acid and is spanned by the “bridge helix”, which may functionally resemble the O-helix found in single-subunit RNAPs (Figure 2) (Cramer, 2002a).

Despite the vast structural differences between single-subunit and multi-subunit RNAPs, numerous mechanistic similarities can be found. Both seem to utilize a similar two-ion dependent catalytic mechanism for chain elongation (Sosunov et al., 2003; Steitz et al., 1994) and both have evolved to neatly accommodate an approximately 8 base pair A-form DNA-RNA hybrid, partially explaining the preference for RNA over DNA as substrate (Kettenberger et al., 2004; Tahirov et al., 2002; Yin and Steitz, 2002). Moreover, the topology of nucleic acid binding appears similar for both types of RNAPs, with the entering DNA duplex being bent by approximately 90° with respect to the exiting DNA-RNA hybrid duplex (Cramer, 2002a; Schwinghammer et al., 2013).

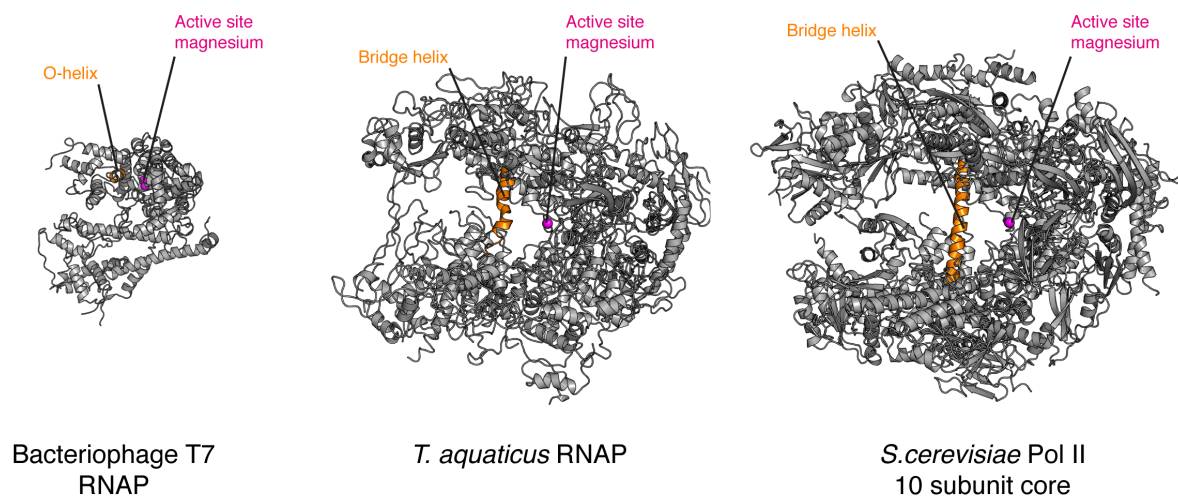


Figure 2. Comparison of single and multi-subunit RNA polymerases.

The structures of T7 RNAP (PDB ID: 1S76; elongation complex with nucleic acid removed) (Yin and Steitz, 2004), *T. aquaticus* RNAP (PDB ID: 1HQM) (Minakhin et al., 2001; Zhang et al., 1999) and the *S. cerevisiae* RNA Pol II 10 subunit core (PDB ID: 1I50) (Cramer et al., 2001) are shown in ribbon representation in grey. The O-helix and bridge helices are shown in orange and active site magnesium ions are shown as magenta spheres.

1.5. Mitochondria

Mitochondria are subcellular, membraneous organelles present in virtually all eukaryotic cells. They are often referred to as the “power houses” of the cell, because the machinery driving ATP synthesis by oxidative phosphorylation as well as numerous key metabolic pathways are compartmentalized in these organelles (Ernster and Schatz, 1981). In addition, mitochondria play important roles in various cellular processes, such as cell signaling, calcium homeostasis, autophagy and apoptosis (Hamanaka and Chandel, 2010; Kroemer and Reed, 2000; Perocchi et al., 2010; Pozzan and Rizzuto, 2000). Consequently, mitochondrial dysfunctions have been implicated in a plethora of human pathologies (Nunnari and Suomalainen, 2012). Furthermore,

due to their high exposure to reactive oxygen species, a byproduct of oxidative phosphorylation, mitochondria are especially prone to molecular damage and this has been implicated in ageing (Nunnari and Suomalainen, 2012).

Phylogenetic analysis suggests that mitochondria have originated from an endosymbiotic event, which marked a decisive step in the evolution of eukaryotic cells (Gray et al., 1999). During this event, an early α -proteobacterium was engulfed by a precursor cell of modern eukaryotes, giving rise to an endosymbiotic relationship that has endured to the present day. This hypothesis is underlined by the finding that mitochondria possess their own genome as well as the entire molecular machinery necessary for the independent expression of genetic information (Hällberg and Larsson, 2014). Apart from the mitochondrial DNA (mtDNA), this includes a full set of transfer RNAs, a mitochondrial RNA polymerase and a dedicated mitochondrial ribosome, which is related to the bacterial ribosome but differs substantially in its protein to RNA ratio (Amunts et al., 2015; Friedman and Nunnari, 2014; Greber and Ban, 2016; Greber et al., 2015). Moreover, mitochondria utilize a distinct genetic code, which requires fewer tRNAs than the universal genetic code (Barrell et al., 1979). The size and organization of the mitochondrial genome is highly divergent, indicating that it evolved in response to varying evolutionary needs in different eukaryotic species (Burger et al., 2003; Gray et al., 1999). Mitochondrial DNA is inherited maternally, as sperm mitochondria are specifically destroyed in the oocyte following fertilization (Sutovsky et al., 1999).

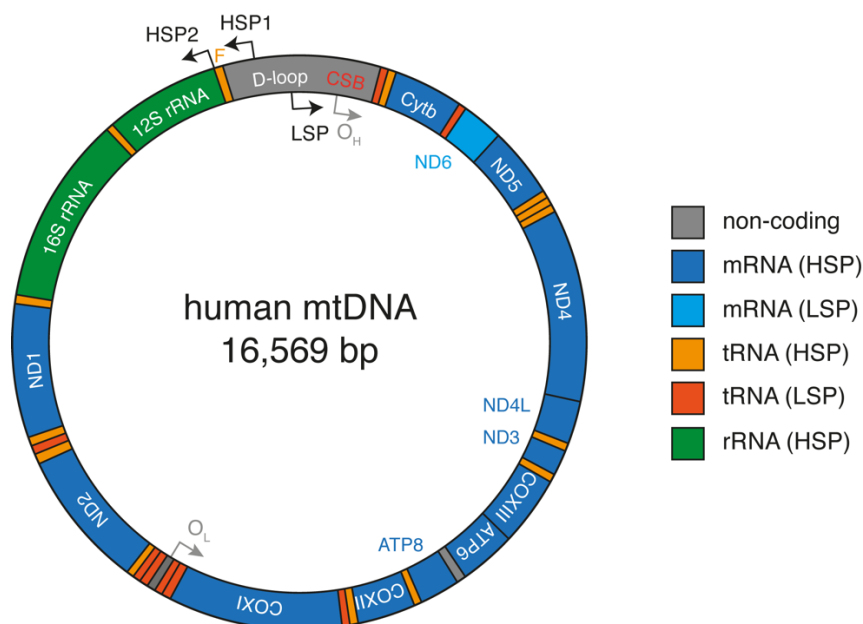


Figure 3. Organization of the human mitochondrial genome.

Schematic depiction of the circular human mitochondrial DNA. Non-coding regions are shown in grey, mRNA coding regions in blue, rRNA coding regions in green and tRNA coding regions in orange. Promoters and origins of replication are indicated. The tRNA^{Phe} gene is labeled with an “F”. LSP: light strand promoter; HSP1/2: heavy strand promoter 1/2; CSB: conserved sequence block; O_{H/L}: origin of replication on the heavy / light strand; Cytb: cytochrome b; ND6: NADH dehydrogenase subunit 6; ND5: NADH dehydrogenase subunit 5; ND4: NADH dehydrogenase subunit 4; COXIII: cytochrome c oxidase subunit III; ATP6: ATP synthase F0 subunit 6; ATP8: ATP synthase F0 subunit 8; COXII: cytochrome c oxidase subunit II; COXI: cytochrome c oxidase subunit I; ND2: NADH dehydrogenase subunit 2; ND1: NADH dehydrogenase subunit 1. Note that the schematic is not drawn to scale. Adapted from (Falkenberg et al., 2007).

In humans, mtDNA is a circular molecule of approximately 16.5 kilo basepairs (kb), encoding for only 13 essential proteins as well as 22 tRNAs and 2 ribosomal RNAs (Figure 3) (Falkenberg et al., 2007). The human mitochondrial proteome, however, is comprised of at least 1500 proteins (Calvo et al., 2016; Mootha et al., 2003; Taylor et al., 2003). Hence, the vast majority of mitochondrial proteins are encoded in the nuclear genome, synthesized in the cytosol and subsequently imported into mitochondria (Neupert, 1997). One hypothesis on the origin of this peculiar situation is that, over the course of evolution, most genes have been transferred from the mtDNA to the nuclear genome (Andersson et al., 2003), leaving only genes in the mitochondrial genome for which synthesis within the organelle is crucial. Strikingly, many of the 13 proteins encoded in human mtDNA are intramembranous components of the respiratory chain and synthesis by a dedicated, specialized ribosome may have been beneficial over the course of evolution (Heijne, 1986).

1.6. Transcription in human mitochondria

The mitochondrial genome is transcribed by a dedicated DNA-dependent RNA polymerase (mtRNAP). Somewhat surprisingly, mtRNAP is not related to multi-subunit RNAPs but to single-subunit bacteriophage RNAPs (Masters et al., 1987; Ringel et al., 2011). However, in contrast to these self-sufficient enzymes, mtRNAP requires additional protein factors for promoter specific initiation, transcript elongation and termination (Gustafsson et al., 2016). Thus, mitochondrial transcription represents an evolutionary “hybrid” between the simple single-subunit RNAPs and the very complex multi-subunit RNAPs. It is interesting to note that not only the mitochondrial transcription machinery, but also the DNA replication machinery seems to be of bacteriophage-origin (Shutt and Gray, 2006).

The individual strands of human mitochondrial DNA can be separated according to their density and are thus designated as the heavy strand and the light strand. Each strand contains one major promoter, termed light strand promoter (LSP) and heavy strand promoter (HSP1), respectively (Figure 3) (Anderson et al., 1981; Montoya et al., 1982; Walberg and Clayton, 1983). It has been proposed that the heavy strand contains a second initiation site, HSP2, located downstream of HSP1 within the tRNA^{Phe} gene (Montoya et al., 1982; 1983). However, the biological relevance of this second promoter is under debate, as transcription from this promoter has been difficult to reconstitute *in-vitro* (Litonin et al., 2010). LSP and HSP drive transcription of polycistronic, near genome-length RNA transcripts, which are further processed by an intricate RNA processing machinery (Aloni and Attardi, 1971a; 1971b; Ojala et al., 1981). Transcription from LSP produces the mRNA for ND6 and eight tRNAs, while transcription from the heavy strand yields 12 mRNAs, both rRNAs and 14 tRNAs (Falkenberg et al., 2007).

1.6.1. Mitochondrial transcription initiation

Human mtRNAP requires two auxiliary proteins, TFAM and TFB2M, to achieve promoter specific transcription initiation (Gustafsson et al., 2016). TFAM, an HMG-Box protein, was the first mammalian mitochondrial transcription factor to be identified (Fisher and Clayton, 1988; Parisi and Clayton, 1991). Biochemical experiments have shown that TFAM binds upstream of

the LSP initiation site and stimulates transcription from both the LSP and the HSP1 promoter in-vitro (Dairaghi et al., 1995a; Litonin et al., 2010). Cross-linking studies suggest a role for TFAM in recruiting mtRNAP to the promoter DNA (Morozov et al., 2014). TFAM has furthermore been demonstrated to bind to DNA non-specifically (Fisher and Clayton, 1988), and it is thus thought to be a packaging factor for mtDNA involved in forming chromatin-like structures called nucleoids (Alam et al., 2003; Chen and Butow, 2005; Fisher et al., 1992). *In vivo* studies have shown that TFAM has a role in regulating mtDNA copy number (Ekstrand et al., 2004) and super-resolution as well as electron microscopic analysis suggest that the rate of compaction of mtDNA is a function of TFAM concentration, and this may be involved in regulating gene expression (Kukat et al., 2015; 2011).

The TFBM family of proteins were initially identified as transcriptional activators based on homology to the yeast protein Mtf1 (Falkenberg et al., 2002; McCulloch et al., 2002). In yeast, the TFAM homolog is not part of the transcriptional machinery and Mtf1 is the only factor required for promoter-specific transcription initiation by the mitochondrial RNA polymerase (Schinkel et al., 1987). In mammals, two paralogous proteins with sequence homology to Mtf1 exist, which are called TFB1M and TFB2M in humans (Falkenberg et al., 2002). Surprisingly, neither Mtf1 nor the mammalian TFBM proteins show significant sequence homology to known proteins involved in transcription. Instead, both show strong sequence similarity to prokaryotic S-adenosylmethionine-dependent methyltransferases that modify ribosomal RNA (Falkenberg et al., 2002; Schubot et al., 2001). *In vitro* studies using recombinant proteins indicate that TFB2M is a transcriptional activator (Litonin et al., 2010), whereas TFB1M seems to function as a methyltransferase, modifying mitochondrial ribosomal RNA at a conserved position (Metodieiev et al., 2009; Seidel-Rogol et al., 2003). Thus, only TFB2M is thought to be a transcription initiation factor in human mitochondria (Litonin et al., 2010). Biochemical experiments have established that it is involved in promoter opening, as mtRNAP can initiate transcription in the absence of TFB2M on a pre-melted DNA template, but not on a fully double-stranded DNA template (Matsunaga and Jaehning, 2004; Posse and Gustafsson, 2016; Sologub et al., 2009). It has been suggested that Mtf1 may interact with the non-template DNA strand in the transcription bubble, mechanistically resembling eukaryotic multi-subunit RNAP initiation factors or the bacterial σ -factor (Paratkar and Patel, 2010). In addition, cross-linking studies indicate that the N-terminus of TFB2M may interact with the priming nucleotide in the initially transcribing complex, suggesting a further role for TFB2M in *de novo* RNA synthesis (Sologub et al., 2009).

Based on biochemical and cross-linking experiments, Morozov et al. have proposed a sequential model for transcription initiation in human mitochondria (Figure 4) (Morozov et al., 2014). According to this model, mtRNAP is recruited to the promoter DNA by TFAM to form the closed pre-initiation complex (preIC). TFB2M then binds to this preIC to form the complete initiation complex (IC), capable of unwinding the duplex DNA around the transcription start site to form an open complex and initiate RNA synthesis (Morozov et al., 2014). Using chemical cross-linking and photo-reactive non-natural amino acids, it was possible to map interactions between the components of the preIC and the IC, respectively. This revealed that the highly divergent N-terminal extension of mtRNAP is the major interaction element with

TFAM (Morozov et al., 2014; 2015). In addition to the previously described interaction of the N-terminus of TFB2M with the priming nucleotide (Sologub et al., 2009), it was proposed that the C-terminal domain of TFB2M may be positioned close to the intercalating hairpin of mtRNAP (Morozov et al., 2015). However, despite these extensive efforts to characterize mitochondrial transcription initiation, both biochemically and by mapping interaction sites, the mechanistic basis of this process remains elusive.

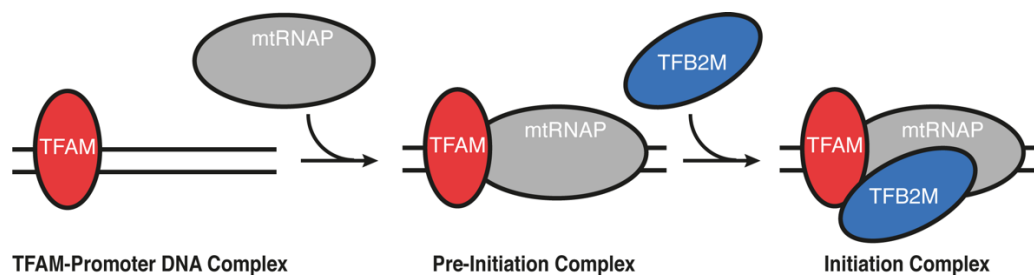


Figure 4. The sequential model of transcription initiation in human mitochondria.

Schematic representation of the steps that lead to transcription initiation. TFAM binds to promoter DNA and recruits mtRNAP to form the closed pre-initiation complex. TFB2M binding leads to formation of the complete initiation complex capable of melting the DNA and initiating RNA synthesis. Adapted from (Morozov et al., 2014)

1.6.2. Transcription and DNA replication are coupled in mitochondria

A further peculiarity of the mitochondrial transcription system is that the mitochondrial RNA polymerase seems to serve a dual function in transcription and in mtDNA replication. This realization emerged when it was found that mtRNAP produces the RNA primer required for initiation of DNA replication (Chang and Clayton, 1985; Kang et al., 1997), somewhat reminiscent of the role of T7 RNA polymerase in replication of the bacteriophage genome (Sugimoto et al., 1987). Mitochondrial DNA contains two major origins of replication, one on each strand, designated O_H and O_L (Figure 3). Several models for the mechanism of mitochondrial DNA replication have been suggested. According to the strand-displacement model, replication of mtDNA is initiated at O_H and proceeds continuously until about two thirds of the genome have been copied and the O_L is exposed. Replication then commences at this second origin and proceeds continuously in the opposite direction until both strands have been faithfully copied (Kang et al., 1997; Robberson et al., 1972; Tapper and Clayton, 1981). Other experiments have suggested that DNA replication in mitochondria may employ conventional, coupled leading- and lagging-strand synthesis, as is the case for nuclear DNA (Holt et al., 2000; Yang et al., 2002). Furthermore, it has been proposed that lagging strand replication may involve transient ribonucleotide incorporation, giving rise to yet another model for mtDNA replication (Yasukawa et al., 2006). As evident from these numerous models, the molecular mechanism of replication in mitochondria remains under debate (Holt and Reyes, 2012). However, recent *in vivo* single-molecule studies strongly support the original strand-displacement model (Phillips et al., 2017). In all models, the RNA primer for initiation of leading strand synthesis at O_H is generated by transcription from the LSP promoter and

premature termination at conserved sequence block II (CSBII) approximately 100 bp downstream of the initiation site (Gillum and Clayton, 1979; Pham et al., 2006). This sequence encodes two consecutive runs of guanines, which have been demonstrated to form a G-quadruplex structure in the nascent RNA and this is thought to be the major determinant for transcription termination (Wanrooij et al., 2010). Furthermore, mtRNAP also produces the RNA primer for lagging strand synthesis originating at O_L (Phillips et al., 2017; Wanrooij et al., 2008). Interestingly, this seems to involve a unique initiation mechanism by mtRNAP at a stem-loop structure in the single-stranded template O_L DNA, without the need for initiation factors (Fusté et al., 2010).

1.6.3. Mitochondrial transcription elongation

The recent discovery of the mitochondrial transcription elongation factor TEFM (Minczuk et al., 2011) revealed a new player in mammalian mitochondrial transcription. TEFM shows only very limited sequence homology to other transcription factors, and this homology is confined to its N-terminal domain (Minczuk et al., 2011). A much more prominent sequence similarity can be found between its larger C-terminal domain and Holliday Junction resolvases, particularly the yeast mitochondrial Holliday Junction resolvase Cce1/Ydc2 (Minczuk et al., 2011). TEFM localizes to mitochondria and was shown to be essential for oxidative phosphorylation activity *in vivo* (Minczuk et al., 2011). Moreover, it was demonstrated to interact with mtRNAP and enhance its processivity as well as raise the abundance of promoter-distal transcripts both *in vivo* and *in vitro* (Minczuk et al., 2011; Posse et al., 2015). In addition to its role as a general elongation factor, TEFM was recently found to have a striking effect on transcription through the CSBII site, where, in the absence of TEFM, mtRNAP terminates transcription to yield the replication primer for O_H (Gillum and Clayton, 1979; Pham et al., 2006). In the presence of TEFM, however, mtRNAP faithfully transcribes through this sequence region, which has been suggested to form extensive secondary structure in the nascent RNA (Agaronyan et al., 2015; Wanrooij et al., 2010). Thus, TEFM seems to be a key player in regulating the switch between productive transcription and mtDNA replication. However, the molecular mechanisms by which TEFM interacts with the elongation complex to enable mtRNAP to transcribe through CSBII and produce long transcripts are not known.

1.7. Previous structural studies of mitochondrial transcription

Over the last decades, the basal mitochondrial transcription machinery has been identified and characterized biochemically. However, it was only over the course of the last years that structural information on the proteins involved in mitochondrial transcription has become available. The first milestone in the structural characterization of this system was the crystal structure of human mitochondrial RNA polymerase, which provided structural evidence for the homology of the catalytic core of mtRNAP to bacteriophage polymerases (Figure 5A and C) (Ringel et al., 2011). In addition, the structure revealed profound differences between T7 RNAP and mtRNAP, providing initial clues for the structural basis of its factor-dependence. Firstly,

mtRNAP contains an N-terminal PPR domain not found in bacteriophage RNAPs. Secondly, the AT-rich recognition loop, a structural element involved in promoter recognition by T7 RNAP, is sequestered in mtRNAP and not available to interact with DNA. Thirdly, the apo form of mtRNAP seems to adopt a conformation in which the major DNA-melting element of T7 RNAP, the intercalating hairpin, is not positioned accordingly to melt DNA (Ringel et al., 2011).

The structure of a transcribing elongation complex (EC), comprised of mtRNAP bound to a DNA/RNA nucleic acid scaffold, offered the first insight into the active conformation of this enzyme (Figure 5B) (Schwinghammer et al., 2013). While the catalytic mechanism of chain elongation seems to be conserved, comparison to T7 RNAP demonstrates further substantial differences between the two polymerases. T7 RNAP undergoes a dramatic re-arrangement of its N-terminal promoter binding domain during the transition from initiation to elongation (Yin and Steitz, 2002), whereas mtRNAP does not refold in a similar fashion (Schwinghammer et al., 2013). Instead, the intercalating hairpin, which is putatively involved in initiation, separates the RNA-DNA hybrid at the upstream edge of the transcription bubble in the mitochondrial EC (Figure 5B) (Schwinghammer et al., 2013).

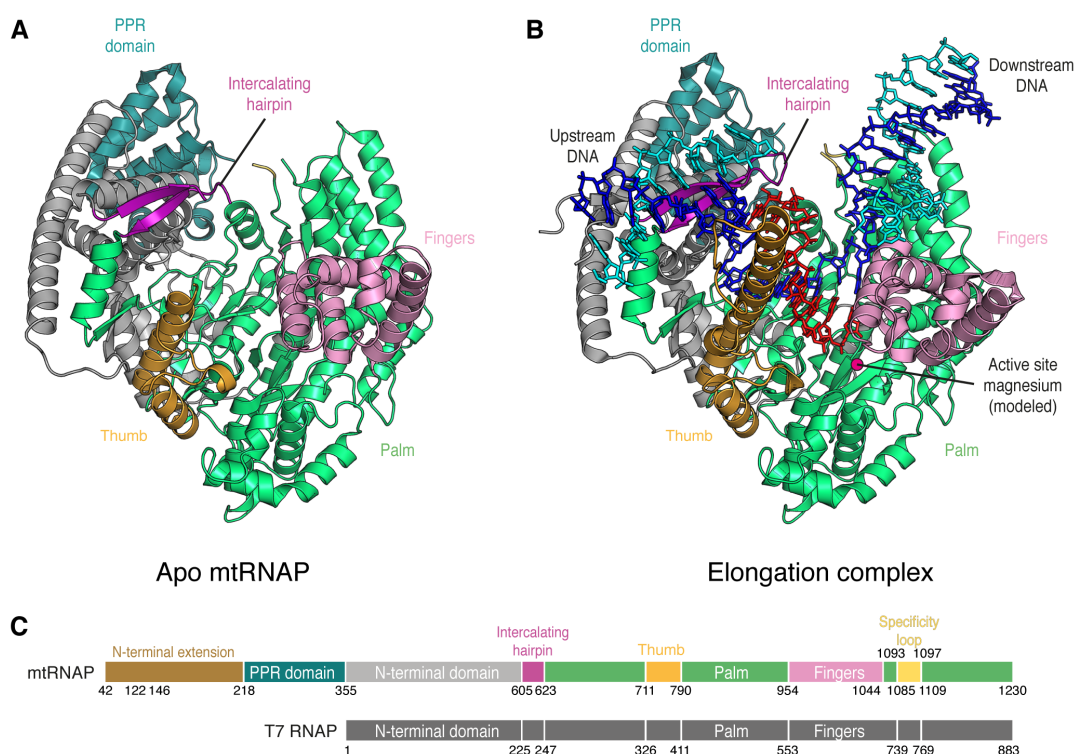


Figure 5. Structures of human mitochondrial RNA polymerase.

(A) Ribbon representation of the structure of free human mtRNAP (PDB ID: 3SPA) (Ringel et al., 2011). Important structural features are indicated. Coloring according to (C). (B) Ribbon representation of the human mitochondrial transcription elongation complex (PDB ID: 4BOC) (Schwinghammer et al., 2013). Depiction as in (A), with nucleic acid depicted as sticks. The non-template strand is colored in cyan, the template strand in blue and the RNA in red. (C) Schematic representation of human mtRNAP and T7 RNAP with important structural elements indicated. Partially adapted from (Ringel et al., 2011).

In 2011, the structure of TFAM in complex with a duplex DNA segment encompassing the LSP binding site was reported by two groups independently (Ngo et al., 2011; Rubio-Cosials et al., 2011). This structure confirmed previous predictions, as it demonstrated that TFAM engages with nucleic acid with both its HMG box domains, each inducing a sharp 90° bend in the bound DNA, resulting in a 180° overall bend (Fisher et al., 1992; Ngo et al., 2011; Rubio-Cosials et al., 2011). The C-terminal tail of TFAM, which is known to be important for transcriptional activation (Dairaghi et al., 1995b), is positioned close to the DNA at the upstream edge of the TFAM binding site. Further structures of TFAM bound to various DNA sequences demonstrated that bending of substrate DNA is not sequence dependent (Ngo et al., 2014). Compared to the LSP promoter-bound structure, the structure of TFAM bound to a part of the HSP promoter sequence suggested a reverse orientation of TFAM with respect to the transcription start site, thus raising a controversy about the architecture of the initiation complex at these two promoters (Ngo et al., 2014). Recent cross-linking experiments, however, argue against different mechanisms of initiation at the different promoters and suggest a conserved architecture for the mitochondrial initiation complex (Morozov and Temiakov, 2016).

Although structures of the yeast TFB2M homolog Mtf1 (Schubot et al., 2001) and mouse TFB1M have been reported (Guja et al., 2013), no structural information on TFB2M is available. These related, isolated structures cannot provide compelling evidence on how Mtf1 or TFB2M may work to promote transcription initiation. In particular, structural information on the interaction between TFB2M, the polymerase and nucleic acid is lacking, thus preventing a mechanistic understanding of how TFB2M assists in transcription initiation.

Taken together, the structural and mechanistic characterization of mitochondrial transcription is still in its infancy compared to the T7 or multi-subunit RNAP transcription systems. Important first steps towards a structural understanding of the mitochondrial transcription cycle have been made by determining the structures of individual proteins and the basal elongation complex. However, structural information on a number of key players and their interaction with the polymerase to form functional complexes is lacking (Figure 6). In particular, structural snapshots of the interplay of initiation factors with mtRNAP and promoter DNA would be needed to decipher the mechanism of factor dependent initiation in mitochondria. Similarly, structural information on TEFM and its interaction with the elongation complex holds promise to unravel the mechanistic basis of how this factor acts to promote processivity and regulate the switch between transcription and replication.

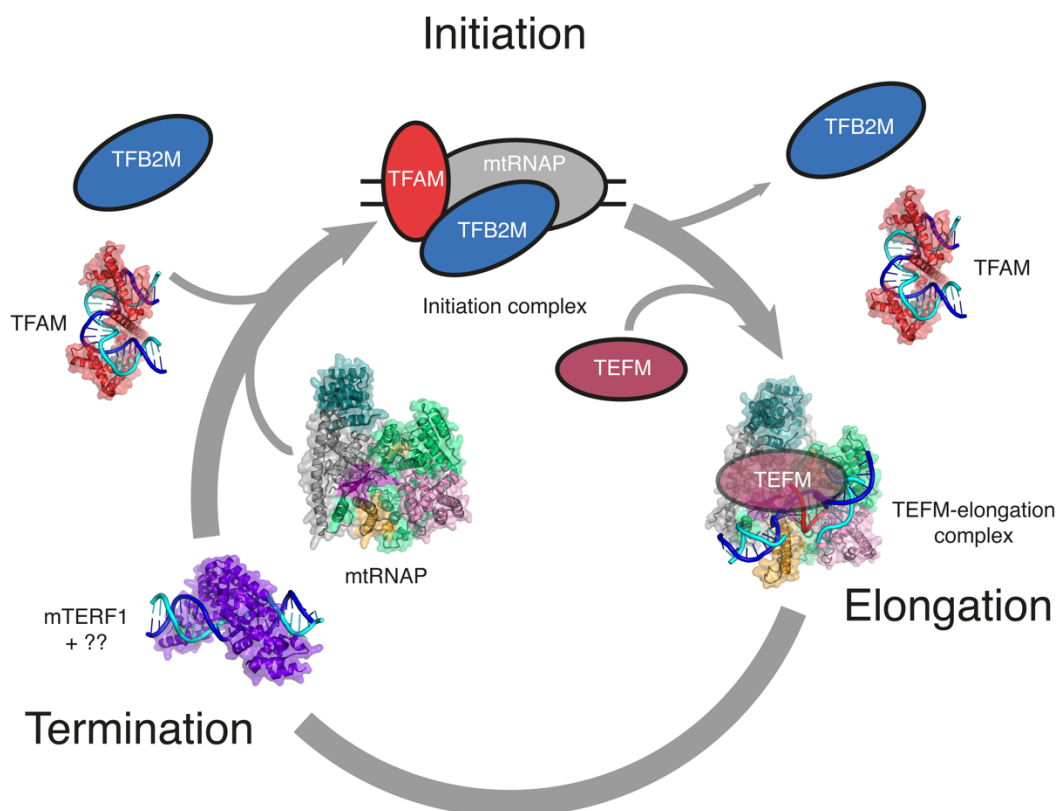


Figure 6. The structural basis of mitochondrial transcription is poorly understood.

Schematic depiction of the transcription cycle in human mitochondria. Known structures are shown in ribbon representation with transparent surfaces, unknown structures are depicted schematically. PDB accession codes of structures used: TFAM: 3TMM (Ngo et al., 2011); mtRNAP: 3SPA (Ringel et al., 2011); mTERF1: 3MVA (Yakubovskaya et al., 2010); EC: 4BOC (Schwinghammer et al., 2013).

1.8. Aim and scope

Despite considerable advances towards understanding transcription and its regulation in human mitochondria, the molecular mechanisms underlying this important process have remained largely enigmatic. This is to a large degree owed to the lack of structural information of the proteins involved and, in particular, of their interplay with each other and nucleic acid during initiation and elongation. Why does mtRNAP require protein factors for all steps of transcription, whereas the related bacteriophage polymerases are self-sufficient? What is the mechanism by which these protein factors, which are evolutionarily unrelated to transcription factors from other systems, act in concert with mtRNAP?

In this study, we attempted to answer these questions by determining the structural basis of transcription initiation and processive elongation.

To decipher the mechanisms governing transcription initiation in human mitochondria, I first determined the structure of human TFB2M at 1.75 Å resolution. This completes our structural picture of the individual components of the basal transcription initiation machinery. In a second step, I then determined the structure of a complete human mitochondrial transcription initiation complex, consisting of mtRNAP, TFAM and TFB2M assembled on a nucleic acid scaffold, at 4.5 Å resolution. This structure reveals how mtRNAP is recruited to promoter-bound TFAM via a novel element in the N-terminal extension of the polymerase, thus positioning it above the transcription start site. It further demonstrates that TFB2M binds near the point of DNA melting, renders mtRNAP capable of promoter melting and stabilizes the melted non-template DNA strand. In collaboration with the Temiakov Lab, we confirmed these findings by biochemical analysis of structure-guided protein mutants.

To unravel the molecular mechanism by which TEFM acts as an elongation factor, I first determined the structure of individual TEFM domains to high resolution (1.3 – 1.9 Å). These data unveil an unusual combination of a Holliday Junction resolvase-like dimerization core with a globular N-terminal domain. In a joint effort with the Temiakov Lab and the Urlaub Lab, we defined the functional roles of the TEFM domains and their interaction with the polymerase and nucleic acids in the elongation complex using transcription assays, nuclease foot-printing and chemical cross-linking. Guided by this data, I crystallized an anti-termination complex comprised of mtRNAP, nucleic acid and the transcriptionally active parts of TEFM. The resulting structure explains how TEFM enhances processivity by stabilizing the open transcription bubble and mobile elements in mtRNAP. Furthermore, it suggests that TEFM prevents formation of a bulky G-quadruplex in the nascent RNA while transcribing through CSBII, thus evading premature transcription termination by mtRNAP.

2. Materials and Methods

This section lists materials and methods used as part of this thesis. Materials and methods performed predominantly by collaborators are not listed here, but can be found in sections 5.1.1, 5.2.1 and 5.2.2.

2.1. Materials

2.1.1. Bacterial strains

Table 1. Bacterial strains used in this study.

Strain	Genotype	Source
BL21-CodonPlus (DE3)-RIL	B F ⁻ <i>ompT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm</i> + Tet ^R <i>gal</i> λ (DE3) <i>endA</i> Hte [<i>argU ileY leuW</i> Cam ^R]	Agilent Technologies
Rosetta2	F ⁻ <i>ompT hsdS</i> _B (r _B ⁻ m _B ⁻) <i>gal dcm</i> pRARE2 (Cam ^R)	Merck Millipore
Rosetta2 pLysS	F ⁻ <i>ompT hsdS</i> _B (r _B ⁻ m _B ⁻) <i>gal dcm</i> (DE3) pLysSRARE2 (Cam ^R)	Merck Millipore
XL-1 Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lacI</i> ^q Δ <i>M15</i> Tn10 (Tet ^R)]	Stratagene

2.1.2. Growth media for *E.coli* and additives

Table 2. Growth media for *E.coli*

Media	Application	Description
LB	Purification of mtRNAP, TFAM, TEFM	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl
SeMet Expression Media	Purification of selenomethionine-labeled proteins	21.6 g/l SelenoMet Base, 5.1 g/l SelenoMet Nutrient Mix, 0.04 mg/ml L-methionine or L-selenomethionine (all from Molecular Dimensions)
ZYP (modified from (Studier, 2005))	Purification of TFB2M	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 50 mM KH ₂ PO ₄ , 50 mM Na ₂ HPO ₄ , 25 mM (NH ₄)SO ₄ , 0.5% (v/v) glycerol, 0.05% (w/v) D-glucose, 2 mM MgSO ₄

Table 3. Additives to growth media

Additive	Used for	Stock concentration	Working concentration
Ampicillin	Resistance selection	100 mg/ml in H ₂ O	100 µg/ml
Chloramphenicol	Resistance selection	30 mg/ml in EtOH	30 µg/ml
D-Glucose	Purification of TEFM	20% (w/v) in H ₂ O	1% (w/v)
IPTG (Isopropyl-β-D-thiogalactopyranoside)	Induction of protein expression	1 M in H ₂ O	0.1 – 0.8 mM
Methionine biosynthesis inhibition mix I	Purification of selenomethionine-labeled proteins	10 mg/ml L-Lysine, 10 mg/ml L-Threonine, 10 mg/ml L-Phenylalanine	0.1 mg/ml L-Lysine, 0.1 mg/ml L-Threonine, 0.1 mg/ml L-Phenylalanine
Methionine biosynthesis inhibition mix II	Purification of selenomethionine-labeled proteins	5 mg/ml L-Leucine, 5 mg/ml L-Isoleucine, 5 mg/ml L-Valine	0.05 mg/ml L-Leucine, 0.05 mg/ml L-Isoleucine, 0.05 mg/ml L-Valine

2.1.3. Plasmids

Table 4. Plasmids used in this study.

Name	Protein variant	Purification tags	Vector	Source
Δ104 mtRNAP tagless	<i>h.sapiens</i> mtRNAP E555A ¹ , lacking residues 1-104	N-terminal 6xHIS, TEV cleavage site	pProEx-Htbmut ²	This study
Δ108 mtRNAP tagless	<i>h.sapiens</i> mtRNAP E555A ¹ , lacking residues 1-108	N-terminal 6xHIS, TEV cleavage site	pProEx-Htbmut ²	This study
Δ119 mtRNAP	<i>h.sapiens</i> mtRNAP E555A ¹ , lacking residues 1-119	N-terminal 6xHIS	pProEx-Htbmut ²	(Morozov et al., 2014)
Δ119 mtRNAP R601E	<i>h.sapiens</i> mtRNAP E555A ¹ , lacking residues 1-119, R601E	N-terminal 6xHIS	pProEx-Htbmut ²	Temiakov lab
Δ42 TFAM “cysless”	<i>h.sapiens</i> TFAM, lacking residues 1-42 and 246, C49S	N-terminal 6xHIS	pET22b	(Sologub et al., 2009)
Δ42 TFAM “cysless” tagless	<i>h.sapiens</i> TFAM, lacking residues 1-42 and 246, C49S	N-terminal 6xHIS, TEV cleavage site	pET22b	This study
Δ20 TFB2M	<i>h.sapiens</i> TFB2M, lacking residues 1-20	N-terminal 6xHIS	pTYB11 Δintein	(Sologub et al., 2009)

Δ20 TFB2M tagless	<i>h.sapiens</i> TFB2M, lacking residues 1-20	N-terminal 6xHIS, TEV cleavage site	pTYB11 Δintein	This study
Δ62 TFB2M	<i>h.sapiens</i> TFB2M, lacking residues 1-62	N-terminal 6xHIS, TEV cleavage site	pTYB11 Δintein	This study
Δ20 TFB2M-Δ268-294-GSSG	<i>h.sapiens</i> TFB2M, lacking residues 1-20; residues 268-294 replaced by GSSG	N-terminal 6xHIS, TEV cleavage site	pTYB11 Δintein	This study
TFB2M ^{cryst}	<i>h.sapiens</i> TFB2M, lacking residues 1-62, residues 268-294 replaced by GSSG	N-terminal 6xHIS, TEV cleavage site	pTYB11 Δintein	This study
Δ20 TFB2M H326A	<i>h.sapiens</i> TFB2M, lacking residues 1-20, H326A	N-terminal 6xHIS	pTYB11 Δintein	Temiaikov lab
Δ20 TFB2M R198A K201A R202A	<i>h.sapiens</i> TFB2M, lacking residues 1-20, R198A, K201A, R202A	N-terminal 6xHIS	pTYB11 Δintein	Temiaikov lab
Δ20 TFB2M Δ8 CTD	<i>h.sapiens</i> TFB2M, lacking residues 1-20 and 388-396	N-terminal 6xHIS	pTYB11 Δintein	Temiaikov lab
Δ20 TFB2M R330A R331A	<i>h.sapiens</i> TFB2M, lacking residues 1-20, R330A, R331A	N-terminal 6xHIS	pTYB11 Δintein	Temiaikov lab
Δ35 TEFM	<i>h.sapiens</i> TEFM, lacking residues 1-35	C-terminal 6xHIS, codon optimized	pET21b	Generated by Kathrin Schwinghammer
Δ50 TEFM	<i>h.sapiens</i> TEFM, lacking residues 1-50	C-terminal 6xHIS, codon optimized	pET21b	Temiaikov lab
Δ135 TEFM	<i>h.sapiens</i> TEFM, lacking residues 1-135	C-terminal 6xHIS, codon optimized	pET21b	Temiaikov lab
Δ144 TEFM	<i>h.sapiens</i> TEFM, lacking residues 1-144	C-terminal 6xHIS, codon optimized	pET21b	This study
Δ159 TEFM	<i>h.sapiens</i> TEFM, lacking residues 1-159	C-terminal 6xHIS, codon optimized	pET21b	This study
Δ50 TEFM-sub140-144	<i>h.sapiens</i> TEFM, lacking residues 1-50, residues 140-144 substituted by AAGAA	C-terminal 6xHIS, codon optimized	pET21b	Temiaikov lab
Δ50 TEFM-sub149-153	<i>h.sapiens</i> TEFM, lacking residues 1-	C-terminal 6xHIS, codon optimized	pET21b	Temiaikov lab

	50, residues 149-153 substituted by AAGAA			
Δ50 TEFM R152A R153A	<i>h.sapiens</i> TEFM, lacking residues 1-50, R152A R153A	C-terminal 6xHIS, codon optimized	pET21b	Temiakov lab

¹ Compared to UniProt sequence (accession code O00411); natural variant.

² pProEx plasmid backbone with mutation in Htb site.

2.1.4. Primers

Table 5. Primers used in this study.

Name	Sequence (5' to 3')	Used to clone	Type
HH 25	TTA ACT TTA AGA AGG AGA TAT ACA TAT GAG TCC GGA AAA TCG TTT TCT GCG	pET21b-Δ144TEFM	Site-directed mutagenesis
HH 26	CGC AGA AAA CGA TTT TCC GGA CTC ATA TGT ATA TCT CCT TCT TAA AGT TAA	pET21b-Δ144TEFM	Site-directed mutagenesis
HH 29	TTA ACT TTA AGA AGG AGA TAT ACA TAT GGA ACG TGA ACG TCT GAA AGC CG	pET21b-Δ149TEFM	Site-directed mutagenesis
HH 30	CGG CTT TCA GAC GTT CAC GTT CCA TAT GTA TAT CTC CTT CTT AAA GTT AA	pET21b-Δ149TEFM	Site-directed mutagenesis
HH 46	GTC ATC ACC ATC ACC ATC ACG AAA ACC TGT ACT TCC AAT CCA ATG CAC GGT TTT GCA TTT TAG GGT C	pTYB11-Δintein-His6-TEV-Δ20TFB2M	Site-directed mutagenesis
HH 47	GAC CCT AAA ATG CAA AAC CGT GCA TTG GAT TGG AAG TAC AGG TTT TCG TGA TGG TGA TGG TGA TGA C	pTYB11-Δintein-His6-TEV-Δ20TFB2M	Site-directed mutagenesis
HH 58	CGA AAA CCT GTA CTT CCA ATC CAA TGC AAA GGC GTC TAA GGC CAG CTT AGA CTT TAA GCG	pTYB11-Δintein-His6-TEV-Δ62TFB2M	Site-directed mutagenesis
HH 59	CGC TTA AAG TCT AAG CTG GCC TTA GAC GCC TTT GCA TTG GAT TGG AAG TAC AGG TTT TCG	pTYB11-Δintein-His6-TEV-Δ62TFB2M	Site-directed mutagenesis
HH 74	GCA TCA CCA TCA CCA TCA CGA AAA CCT GTA CTT CCA ATC CAA TGC ATC ATC TGT CTT GGC AAG TAG	pET22b-His6-TEV-Δ42TFAM-cysless	Site-directed mutagenesis
HH 75	TGC ATT GGA TTG GAA GTA CAG GTT TTC GTG ATG GTG ATG GTG ATG CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC	pET22b-His6-TEV-Δ42TFAM-cysless	Site-directed mutagenesis
HH 99 ¹	ATG GGG GCC AAG GAT GCC ACC CCG GTG CCC TGT GGC CGC TGG GCA AAG ATA CTG G	pProEx-Htbmut-His6-TEV-Δ104mtRNAP	Around the Horn
HH 100 ¹	AGG AAG GTC CAG ATG GGG GCC AAG GAT GCC ACC CCG GTG CCC	pProEx-Htbmut-His6-TEV-Δ104mtRNAP	Around the Horn

	TGT GGC CGC TGG GCA AAG ATA CTG G		
HH 101 ¹	TGC ATT GGA TTG GAA GTA CAG GTT TTC GTG ATG	pProEx-Htbmut-His6- TEV-Δ104mtRNAP, pProEx-Htbmut-His6- TEV-Δ108mtRNAP	Around the Horn
HH 102 ¹	GAT ATG GAA GCA AGC CTG TAT GCC GAT CTG AAT AAA ACC TTT GCA CAG G	pET21b-Δ35TEFM- F244E-L248D-I252D- M256S-L260D	Around the Horn
HH 103 ¹	ATG AAA GTG ATC CAG AAT CGG TTC CAG GCT GCT ATT CTG AAT GC	pET21b-Δ35TEFM- F244E-L248D-I252D- M256S-L260D	Around the Horn
HH 133 ¹	AAG CTG TAT CTT ATT CAA ATG ATT CCT CGT C	pTYB11-Δintein-His6- TEV-Δ62TFB2M-Δ268- 294-GSSG, pTYB11- Δintein-His6-TEV- Δ20TFB2M-Δ268-294- GSSG	Around the Horn
HH 135 ¹	ACC GGA GGA GCC AGG CTC CAT GTG CAG AAC C	pTYB11-Δintein-His6- TEV-Δ62TFB2M-Δ268- 294-GSSG, pTYB11- Δintein-His6-TEV- Δ20TFB2M-Δ268-294- GSSG	Around the Horn
T7 promoter	TAA TAC GAC TCA CTA TAG GGG	-	sequencing
T7 terminator	CTA GTT ATT GCT CAG CGG TG	-	sequencing
M13 pUC rev	AGC GGA TAA CAA TTT CAC ACA GG	-	sequencing
pTRCHis rev	CTG ATT TAA TCT GTA TCA GG	-	sequencing
mtRNAP seq1 for	ACT GTT CTG AAG GCC GTG C	-	sequencing
mtRNAP seq2 for	AGA TGC CAT GCA GCC TGG	-	sequencing
mtRNAP seq3 for	AGT TCG CCC AGG GCC GCC	-	sequencing
mtRNAP seq4 for	TTC AGA GCA TCA CCT ACA CCC	-	sequencing
mtRNAP seq5 rev	AGG GTC TTC AGG GTC TTC C	-	sequencing

¹Synthesized with 5'-phosphate group.

2.1.5. Synthetic oligonucleotides

Table 6. Synthetic oligonucleotides used for reconstitution of transcription complexes.

Name	Sequence (5' to 3') ¹	Application	Source
HH NT2	TGT TAG TTG GGG GGT GAC TGT TAA AAG TGC ATA CCT ATC CCC GAT AG GCC	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH TS1	GGC CTA TCT TTT GGC GGT ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH NT6	CAC CGC TGC TAA CCC CAT ACC CCG AAC CAA CCA AAT TAT CCC GAC AGG CC	In-vitro reconstitution of HSP initiation complex for crystallization	This study
HH TS3	GGC CTG TCT TTG GGG TTT GGT TGG TTC GGG GTA TGG GGT TAG CAG CGG TG	In-vitro reconstitution of HSP initiation complex for crystallization	This study
HH TS1- Bromine1	GGC C-5Br-dU-A 5Br-dU- CT TTT GGC GGT ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH NT2- Bromine2	TGT TAG TTG GGG GGT GAC TGT TAA AAG TGC ATA CC-5- Br-dU ATC CCC GAT AG-5-Br- dU CC	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH TS50	GGA CTA TCT TTT GGC GGT ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH TS1- Bromine3	GGC CTA TCT TTT GGC GG-5- Br-dU ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH NT- Bromine4	TGT TAG TTG GGG GGT GAC TGT TAA AAG 5-Br-dU-GC ATA CCT ATC CCC GAT AGG CC	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH TS1- Bromine5	GGC CTA TCT TTT GG-5Br-dU GGT ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH NT67	TGT TAG TTG GGG GGT GAC TGT TAA AAG TGC ATA CAT ATC CCC GAT AGG CC	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH TS- Bromine6	GGC CTA TCT TTT GGC 5Br- dU-GT ATG CAC TTT TAA CAG TCA CCC CCC AAC TAA CA	In-vitro reconstitution of LSP initiation complex for crystallization	This study
HH NT25	GAA CAT GGT GTA ATT ATT TCG ACG CCA GAC GAA C	In-vitro reconstitution of EC- TEFM complex for crystallization	This study
HH TS22	GTT CGT CTG GCG TGC GCG CCG CTA CAC CAT GTT C	In-vitro reconstitution of EC- TEFM complex for crystallization	This study

HH RNA16	agu cug cgg cgc gc	In-vitro reconstitution of EC-TEFM complex for crystallization	(Schwingha mmer et al., 2013)
HH NT19	CAT GGG GTA ACT AGT TCG ACG CCA GAC G	In-vitro reconstitution of EC-TEFM complex for BS3 / EDC crosslinking	This study
HH TS16	CGT CTG GCG TGA TCA CGA CTA CCC CAT G	In-vitro reconstitution of EC-TEFM complex for BS3 / EDC crosslinking	This study
HH RNA14	uga ugg uaa ugc ucc ugu cgu gau c	In-vitro reconstitution of EC-TEFM complex for BS3 / EDC crosslinking	This study

[†] 5-Br-dU = 5-Bromo-Uracil;

UPPERCASE = DNA; lowercase = RNA

2.1.6. Buffers and solutions

Table 7. General solutions

Name	Description or Source	Application
100x Protease Inhibitor Mix	0.028 mg/ml Leupeptin, 0.137 mg/ml Pepstatin A, 0.017 mg/ml PMSF, 0.33 mg/ml benzamidine in 100% EtOH (p.a.)	Protein purification
10x TAE	2.5 M Tris-acetate, 50 mM EDTA pH 8.0	Agarose gels
20 x MES SDS running buffer	1M MES pH 7.3, 1M Tris Base, 2% SDS, 20 mM EDTA (Thermo Fisher Scientific)	SDS-PAGE
20 x MOPS SDS running buffer	1M MOPS pH 7.7, 1M Tris Base, 2% SDS, 20 mM EDTA (Thermo Fisher Scientific)	
5x SDS-PAGE sample buffer	250 mM Tris-HCl pH 6.8 at 20 °C, 50% (v/v) glycerol, 0.03% (w/v) bromophenol blue, 7.5% (w/v) SDS, 100 mM DTT	SDS-PAGE
Gel staining solution	InstantBlue, Expedition	SDS-PAGE visualization
NuPAGE 4x LDS Sample Buffer	Invitrogen	SDS-PAGE
PageRuler Prestained Protein Ladder, 10 to 180 kDa	Thermo Fisher Scientific	SDS-PAGE
PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na ₂ HPO ₄ , 1.8 mM KH ₂ PO ₄	various
Protein gel electrophoresis running buffer	NuPAGE MES or MOPS SDS running buffer (Invitrogen)	SDS-PAGE

Table 8. Buffers used for purification of mtRNAP

Name	Description
mtRNAP lysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 10 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
mtRNAP Ni-NTA wash buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 1.5 M NaCl, 20 mM MgCl ₂ , 10 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
mtRNAP Ni-NTA elution buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 250 mM imidazole, 5% (v/v) glycerol, 2 mM DTT
mtRNAP dialysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer A	40 mM Tris-HCl pH 8.0 at 23 °C, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer B	40 mM Tris-HCl pH 8.0 at 23 °C, 2 M NaCl, 5% (v/v) glycerol, 2 mM DTT
mtRNAP size exclusion buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 5% (v/v) glycerol, 10 mM DTT

Table 9. Buffers used for purification of TFAM

Name	Description
TFAM lysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 2 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TFAM Ni-NTA wash buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 1.5 M NaCl, 2 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TFAM Ni-NTA elution buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 250 mM imidazole, 5% (v/v) glycerol, 2 mM DTT
TFAM dialysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer A	40 mM Tris-HCl pH 8.0 at 23 °C, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer B	40 mM Tris-HCl pH 8.0 at 23 °C, 2 M NaCl, 5% (v/v) glycerol, 2 mM DTT
TFAM size exclusion buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 100 mM NaCl, 5% (v/v) glycerol, 10 mM DTT

Table 10. Buffers used for purification of TFB2M

Name	Description
TFB2M lysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 20 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TFB2M Ni-NTA wash buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 1.5 M NaCl, 20 mM MgCl ₂ , 20 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TFB2M Ni-NTA elution buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 250 mM imidazole, 5% (v/v) glycerol, 2 mM DTT

TFB2M dialysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer A	40 mM Tris-HCl pH 8.0 at 23 °C, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer B	40 mM Tris-HCl pH 8.0 at 23 °C, 2 M NaCl, 5% (v/v) glycerol, 2 mM DTT
TFB2M size exclusion buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 100 mM NaCl, 5% (v/v) glycerol, 10 mM DTT

Table 11. Buffers used for purification of TEFM

Name	Description
TEFM lysis buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 20 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TEFM Ni-NTA wash buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 1.5 M NaCl, 20 mM MgCl ₂ , 20 mM imidazole, 5% (v/v) glycerol, 2 mM DTT, 1 x PI
TEFM Ni-NTA elution buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 300 mM NaCl, 20 mM MgCl ₂ , 250 mM imidazole, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer A	40 mM Tris-HCl pH 8.0 at 23 °C, 5% (v/v) glycerol, 2 mM DTT
Ion exchange buffer B	40 mM Tris-HCl pH 8.0 at 23 °C, 2 M NaCl, 5% (v/v) glycerol, 2 mM DTT
TEFM size exclusion buffer	40 mM Tris-HCl pH 8.0 at 23 °C, 100 mM NaCl, 5% (v/v) glycerol, 10 mM DTT

2.2. Methods

2.2.1. General methods

Site-directed mutagenesis PCR

QuikChange PCR was carried out in 20 µl reactions containing Phusion High Fidelity Polymerase (homemade), Phusion High Fidelity Buffer (New England Biolabs), 0.2 mM dNTPs, 0.05 µM of each primer and 2.5 ng of template DNA. The reaction was carried out for 25 cycles each consisting of a 1 min denaturation step at 95 °C, followed by an annealing step at 55 °C for 1 min and an elongation step at 72 °C, for which the time was adjusted to the length of the template. Following thermocycling, 1 µl of DpnI (New England Biolabs) was added and the reaction was incubated at 37 °C for at least 1 h prior to transformation of the entire reaction into competent *E.coli* XL1-Blue cells.

Around-the-Horn PCR

For „Around-the-Horn“ PCR, primer pairs either containing an insertion or generating a deletion were designed to anneal in opposite orientations to the respective template sequence and purchased as chemically synthesized, 5'-phosphate containing oligonucleotides. PCR reactions were typically performed in Q5 Polymerase reaction buffer (New England Biolabs) with 0.5 μ M of each primer, 0.2-2.0 ng of template DNA, 0.2 mM of each dNTP and 20 units of Q5 High-Fidelity DNA Polymerase (New England Biolabs). Reaction conditions (annealing temperature and extension time) were adjusted to the respective template and PCR reactions were carried out for 35 cycles. Template DNA was subsequently digested by addition of 1 μ l DpnI (20 units/ μ l, New England Biolabs) and incubation for at least 1 h at 37 °C. The reaction products were analyzed by electrophoresis in 0.5 – 1.5% agarose gels containing Sybr Safe® (Thermo Fisher) and run in 1 x TAE buffer at 100 V. Bands corresponding to PCR products were excised, purified using a Gel Extraction Kit (Qiagen) and eluted in 25 μ l ddH₂O. The linear reaction products were ligated by addition of 3 μ l T4 DNA Ligase Buffer (New England Biolabs) and 2 μ l T4 DNA Ligase (400 units/ μ l, New England Biolabs) and incubation at 16°C for 2 h.

Transformation of *E.coli* cells

For transformation, 50 μ l aliquots of chemically competent *E.coli* cells were incubated with either purified plasmid DNA (50 – 200 ng) or PCR products (entire reaction) on ice for 20 min. Cells were heat-shocked at 42 °C for 45 s and immediately placed on ice again for 5 min. The cells were recovered by addition of 1 ml of LB media and incubation at 37 °C and 600 rpm for 1 h. The cells were pelleted by centrifugation at 3000 rpm in a table top centrifuge for 5 min and resuspended in 100 μ l of fresh LB media. The cells were plated on LB-agar plates containing the respective antibiotics for selection and incubated at 37 °C overnight.

Isolation of plasmid DNA and sequence verification

Single colonies were picked from agar plates and used to inoculate 6 ml of LB media containing the appropriate antibiotics for selection. The culture was incubated at 37 °C and 160 rpm overnight. Plasmid purification was performed using a Mini Prep Kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA was eluted in ddH₂O and stored at -20 °C. Sequences were verified by sequencing (Seqlab) using appropriate primers.

Preparation of glycerol stocks for protein expression

Following transformation of the respective expression plasmid into chemically competent *E.coli* cells and plating on LB agar plates containing the appropriate antibiotics for selection, an initial overnight culture of LB media containing the appropriate antibiotics was inoculated from multiple colonies and incubated at 37 °C and 160 rpm overnight. A glycerol stock was prepared by mixing an equal volume of the overnight culture and 50% (v/v) glycerol (25% (v/v)

final concentration) and stored at -80 °C. Glycerol stocks were used to inoculate subsequent overnight cultures for large scale protein expression.

Protein and nucleic acid concentration determination

Concentrations for purified proteins and protein complexes were determined using a NanoDrop spectrophotometer (Thermo Fisher) using theoretical extinction coefficients calculated using the ExPASy ProtParam server (Wilkins et al., 1999) (for proteins) or the IDT UV spectrum analyzer (<http://biophysics.idtdna.com/UVSpectrum.html>) (for nucleic acid).

Protein identification by mass spectrometry

Protein identification by mass spectrometry was performed in the laboratory of Henning Urlaub at the Max-Planck-Institute for Biophysical Chemistry in Göttingen, Germany.

Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE analysis was used to monitor progress of protein purifications and estimate the purity of protein preparations. Furthermore, it was used to estimate stoichiometry of *in-vitro* reconstituted protein complexes. For analysis, protein samples were mixed with either 5 x SDS-PAGE sample buffer (homemade) or 4 x LDS sample buffer (Thermo Fisher) (Table 7) and boiled at 95 °C for 3-5 min. The samples were spun down, loaded on a NuPage 4 – 12% BIS-TRIS gel (Thermo Fisher) together with 5 µl of PageRuler Prestained Protein Ladder (Thermo Fisher) and separated by electrophoresis at 200 V in either NuPage MES or MOPS Running Buffer (Thermo Fisher). Bands were visualized using InstantBlue stain (Expedeon).

2.2.2. Protein purification

Purification of human mitochondrial RNA polymerase

Human mtRNAP was expressed in *E.coli* BL21 (DE3) RIL cells (Agilent) and purified using an optimized protocol based on the previously published purification (Sologub et al., 2009). Cells were grown at 37 °C and 160 rpm until the OD₆₀₀ reached 0.6 units and expression was induced by addition of 0.15 mM IPTG and carried out at 16 °C for 18 h. The cells were harvested by centrifugation at 7,800 g for 20min and the cell pellets were flash frozen in liquid nitrogen and stored at -80 °C. For purification, cell pellets were thawed and resuspended in cold mtRNAP lysis buffer. The cells were lysed by ultrasonification (0.4 s pulses on, 0.6 s off, 60% amplitude, 10 min total time) and cell debris was pelleted by centrifugation at 57,900 g for 30 min twice. 20 ml of nickel-nitrilotriacetic acid agarose (Ni-NTA beads, Qiagen) equilibrated with mtRNAP lysis buffer were added to the supernatant and incubated on a rotating wheel at

4 °C for 1.5 h. The beads were washed twice with 10 bead volumes of mtRNAP Ni-NTA wash buffer for 30 min and subsequently with 10 bead volumes of lysis buffer to remove high salt. Bound protein was eluted in 5 bead volumes of mtRNAP Ni-NTA elution buffer. Protein variants bearing a non-cleavable purification tag were immediately subjected to heparin affinity chromatography. For constructs containing a TEV-cleavable 6xHIS-tag, TEV protease (2-4 mg, homemade) was added and the sample was dialyzed overnight at 4 °C against mtRNAP dialysis buffer using SnakeSkin® dialysis tubing (7 kDa cut-off, Thermo Fisher). Heparin affinity chromatography was carried out using multiple HiTrap HP 5ml Heparin columns (GE Healthcare) equilibrated with a buffer containing 85% (v/v) ion exchange buffer A and 15% (v/v) ion exchange buffer B. The protein was eluted with a linear gradient from 15 to 50% ion exchange buffer B over 15 CV. Peak fractions corresponding to mtRNAP were collected and concentrated in a centrifugal concentrator (30 kDa cut-off, Amicon). The protein was then subjected to size exclusion chromatography using a Superdex 200 10/300 Increase column (GE Healthcare) equilibrated with mtRNAP size exclusion buffer. Peak fractions corresponding to mtRNAP were pooled and concentrated in a centrifugal concentrator (30 kDa cut-off, Amicon) to a final concentration of 8-10 mg/ml. The protein was aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C.

Purification of human TFAM

Human TFAM was expressed in *E.coli* BL21 Rosetta2 pLysS cells (Merck Millipore) and purified using an optimized protocol based on the previously published purification (Sologub et al., 2009). Cells were grown at 37 °C and 160 rpm until the OD₆₀₀ reached 0.8-1.0 units and expression was induced by addition of 0.8 mM IPTG and carried out at 37 °C for 1.5 h. The cells were harvested by centrifugation at 7,800 g for 20min and the cell pellets were flash frozen in liquid nitrogen and stored at -80 °C. For purification, cell pellets were thawed and resuspended in cold TFAM lysis buffer. The cells were lysed by ultrasonification (0.4 s pulses on, 0.6 s off, 60% amplitude, 15 min total time) and cell debris was pelleted by centrifugation at 57,900 g for 30 min twice. The supernatant was loaded on a HisTrap HP 5ml column (GE Healthcare) equilibrated with TFAM lysis buffer and the column was washed with 40 CV of TFAM Ni-NTA wash buffer followed by 20 CV of TFAM lysis buffer to remove excess salt. TFAM was eluted with 10 CV of TFAM Ni-NTA elution buffer. Protein variants bearing a non-cleavable purification tag were immediately subjected to heparin affinity chromatography. For constructs containing a TEV-cleavable 6xHIS-tag, TEV protease (4 mg, homemade) was added to cleave the purification tag and the sample was dialyzed overnight against TFAM dialysis buffer using SnakeSkin® dialysis tubing (7000 kDa cut-off, Thermo Fisher). To remove un-cleaved protein and the 6xHIS-tagged TEV protease, the sample was subjected to a second Ni-NTA affinity chromatography step. For this, the sample was supplemented with imidazole to approx. 60 mM final concentration and applied to a HisTrap HP 5 ml column (GE Healthcare). The flow-through was collected and the column was washed with 5 CV of TFAM dialysis buffer. For heparin affinity chromatography, the flow-through and wash fractions were combined and applied to multiple HiTrap Heparin HP 5 ml columns (GE Healthcare) equilibrated with a buffer consisting of 85% (v/v) ion exchange buffer A and 15% (v/v) ion

exchange buffer B. The protein was eluted with a linear gradient from 15 to 50% ion exchange buffer B over 10 CV. Peak fractions corresponding to TFAM were pooled, diluted with ion exchange buffer A to approx. 300 mM NaCl and subjected to cation exchange chromatography on HiTrap SP HP 5ml columns (GE Healthcare) equilibrated with a buffer containing 85% (v/v) ion exchange buffer A and 15% (v/v) ion exchange buffer B. Bound protein was eluted with a linear gradient from 15 to 50% ion exchange buffer B over 15 CV. Peak fractions corresponding to TFAM were pooled and subjected to a third Ni-NTA step at low imidazole concentration to remove trace amounts of TFAM bearing the tag, which were still present as judged by SDS-PAGE. For this, the sample was supplemented with 2 mM imidazole and applied to a HisTrap HP 5 ml column (GE Healthcare) equilibrated with a buffer consisting of 80% (v/v) ion exchange buffer A and 20% (v/v) ion exchange buffer B. The column was washed with 3-5 CV of buffer and the flow-through and wash fractions were combined and concentrated in a centrifugal concentrator (3 kDa cut-off, Amicon). The sample was further purified by size exclusion chromatography on a HiPrep 16/600 Superdex 75 pg column (GE Healthcare) equilibrated with TFAM size exclusion buffer. Peak fractions corresponding to TFAM were pooled and concentrated in a centrifugal concentrator (3 kDa cut-off, Amicon) to 20-25 mg/ml final concentration. TFAM was aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C.

Purification of human TFB2M

Human TFB2M was expressed in *E.coli* BL21 (DE3) RIL cells (Agilent) and purified using an optimized protocol based on the previously published purification (Sologub et al., 2009). Cells were grown in modified ZYP media (see Table 2) at 37 °C and 160 rpm until the OD₆₀₀ reached 6.0 units. Expression was induced by addition of 0.1 mM IPTG and carried out at 16 °C for 18 h. The cells were harvested by centrifugation at 7,800 g for 20min and the cell pellets were flash frozen in liquid nitrogen and stored at -80 °C. For purification, cell pellets were thawed and resuspended in cold TFB2M lysis buffer. The cells were lysed by ultrasonification (0.4 s pulses on, 0.6 s off, 75% amplitude, 15 min total time) and cell debris was pelleted by centrifugation at 57,900 g for 30 min twice. 20 ml of Ni-NTA agarose beads (Qiagen) equilibrated with TFB2M lysis buffer were added to the supernatant and incubated on a rotating wheel at 4 °C for 1.5 h. The beads were washed with 10 times the bead volumes of TFB2M Ni-NTA wash buffer 30 min followed by 10 bead volumes of lysis buffer to remove high salt. The bound protein was eluted in 2.5 bead volumes of TFB2M Ni-NTA elution buffer. Protein variants bearing a non-cleavable purification tag were immediately subjected to heparin affinity chromatography. For constructs containing a TEV-cleavable 6xHIS-tag, TEV protease (2 mg, homemade) was added and the sample was dialyzed against TFB2M dialysis buffer overnight using SnakeSkin® dialysis tubing (7 kDa cut-off, Thermo Fisher). To remove un-cleaved protein and the 6xHIS-tagged TEV protease, the sample was subjected to a second Ni-NTA affinity chromatography step. For this, the sample was supplemented with imidazole to approx. 60 mM final concentration and applied to a HisTrap HP 5 ml column (GE Healthcare). The flow-through was collected and the column was washed with 5 CV of TFAM dialysis buffer. The flow-through and wash fractions were combined and subjected to heparin affinity chromatography using multiple HiTrap Heparin HP 5ml columns equilibrated with a buffer

consisting of 85% (v/v) ion exchange buffer A and 15% (v/v) ion exchange buffer B. The bound protein was eluted using a linear gradient from 15% (v/v) to 50% (v/v) over 15 CV. Peak fractions containing TFB2M were pooled and concentrated in a centrifugal concentrator (10 kDa cut-off, Amicon). The sample was then subjected to size exclusion chromatography using a Superdex 200 10/300 Increase column (GE Healthcare) equilibrated with TFB2M size exclusion buffer. Peak fractions containing TFB2M were pooled, concentrated in a centrifugal concentrator (10 kDa cut-off, Amicon) to 20 mg/ml. Purified TFB2M was aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C.

Purification of human TEFM

Human TEFM was expressed in *E.coli* BL21 (DE3) RIL (Agilent) or Rosetta2 (DE3) cells (Merck Millipore). Cells were grown in LB media containing 1% D-glucose at 37 °C and 160 rpm until the OD₆₀₀ reached 0.6 units. Expression was induced by addition of 0.15 mM IPTG and carried out at 16 °C for 18 h. The cells were harvested by centrifugation at 7,800 g for 20 min and the cell pellets were flash frozen in liquid nitrogen and stored at -80 °C. For purification, cell pellets were thawed and resuspended in cold TEFM lysis buffer. The cells were lysed by ultrasonification (0.4 s pulses on, 0.6 s off, 60% amplitude, 10 min total time) and the cell debris was pelleted by centrifugation at 57,900 g for 30 min twice. The supernatant was loaded on a HisTrap HP 5ml column (GE Healthcare) equilibrated with TEFM lysis buffer and the column was washed with 10 CV of TEFM Ni-NTA wash buffer followed by 10 CV of TEFM lysis buffer to remove excess salt. TEFM was eluted in 10 CV of TEFM Ni-NTA elution buffer and subsequently loaded onto multiple HiTrap Heparin HP 5ml columns (GE Healthcare) equilibrated with a buffer containing 85% (v/v) ion exchange buffer A and 15% (v/v) ion exchange buffer B. The bound protein was eluted with a linear gradient from 15 to 50% ion exchange buffer B over 15 CV. Peak fractions corresponding to TEFM were pooled and concentrated using a centrifugal concentrator (10 kDa cut-off, Amicon). The protein was further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated with TEFM size exclusion buffer. Peak fractions were pooled and concentrated using a centrifugal concentrator (10 kDa cut-off, Amicon) to approx. 20 mg/ml final concentration. TEFM was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

Purification of selenomethionine-labeled proteins

Selenomethionine-labeled proteins were expressed using the methionine biosynthesis inhibition method (Doublié, 2007; Van Duyne et al., 1993) in the same *E.coli* expression strains as used for regular protein expression. An overnight culture of LB media containing the respective antibiotics was grown as described above. The overnight culture was used to inoculate 1 l of SeMet Base media (Molecular Dimensions) (1:100) supplemented with L-methionine (0.04 mg/ml). The culture was grown at 37 °C and 140 rpm until the OD₆₀₀ reached approx. 1.0 units. The cells were pelleted by centrifugation at 7,800 g for 20 min and washed twice with 50 ml of PBS and resuspended in a further 50 ml of PBS. Protein expression was carried out in SeMet

base media (Molecular Dimensions) supplemented with L-selenomethionine (0.04 mg/ml) which was inoculated with the washed cell suspension (1:250). The cells were grown at 37 °C and 140 rpm until the appropriate OD₆₀₀ was reached (in the case of TFB2M, an OD₆₀₀ of 0.75 was used) and protein expression was induced by addition of IPTG as described above. Purification of selenomethionine-labeled proteins was carried out as for the native protein variants, with the DTT concentration increased from 2 mM to 5 mM in all buffers (except for size exclusion buffers) to prevent oxidation of selenomethionine-residues.

2.2.3. General crystallization methods

Microseeding of crystals

To prepare solutions for microseeding, drops containing initial crystals were diluted with 1-5 drop volumes of reservoir solution and mixed by pipetting up and down until the crystals were visibly crushed. The entire solution was then transferred to a Seed Bead kit (Hampton) and further diluted with reservoir solution to a final volume of 50 µl. The Seed Bead tube was vortexed in 30 s intervals for a total of 3 min, with incubation on ice in between each interval to prevent heating of the sample. The Seed Bead tube was subsequently sonicated in a water bath at 4 °C for 2 min. A ten-fold dilution series was prepared from the seed stock using reservoir solution. Crystals were typically microseeded by setting up new drops consisting of equal volumes of protein solution and reservoir solution to which 0.2–0.5 µl of seeding solution were added.

Cryo-protection and freezing of crystals

Crystals of the initiation complex and the EC-TEFM complex were cryo-protected by gradually increasing the glycerol concentration in the drop. For this, three solutions resembling the respective reservoir solution of the drop with increasing glycerol concentrations were prepared (6%, 12.5%, 25%, (v/v)). One drop volume of the first solution was added to the drop containing crystals and mixed by pipetting until the drop appeared homogenous under the microscope. One drop volume was then removed from the drop again. This process was repeated three times for each glycerol concentration step. For the final step, as much solution as possible was removed from the drop without harming the crystals and the drop was then entirely replaced by the 25% glycerol solution. Crystals of the TEFM CTD were cryo-protected in a similar fashion, but in a 2-step process using reservoir solutions containing 10% and 20% glycerol, respectively. Crystals of TFB2M were cryo-protected by increasing the glycerol concentration to 25% (v/v) using a single drop-exchange protocol. Following cryo-protection, crystals were fished and flash frozen in liquid nitrogen.

2.3. Structural basis of mitochondrial transcription initiation

This section contains methods specifically related to the study on the structure of the human mitochondrial transcription initiation complex. Only those methods developed and performed as part of this thesis are described here, while methods carried out predominantly by the collaborators from the Temiakov lab are described in section 5.1.1. Parts of the methods described here are currently in the process of peer review for publication:

H.S. Hillen, Y.I. Morozov, A. Sarfallah, D. Temiakov and P. Cramer (2017) Structural basis of mitochondrial transcription initiation. *Cell*, in revision

A detailed list of author contributions can be found on page VI.

Protein expression constructs

A construct of human mtRNAP lacking the N-terminal 104 amino acids (Δ 104 mtRNAP) with an N-terminal 6xHIS tag and a TEV cleavage site was constructed from the pProEx-based expression plasmid described previously (Sologub et al., 2009). Human TFAM was expressed as a construct lacking the mitochondrial localization sequence (res. 1-42) and two cysteine residues (C49S and without the C-terminal C246), mutations which have been shown to have no effect on activity (Morozov et al., 2014). The construct was cloned from the previously described pET22b-based expression plasmid (Morozov et al., 2014) by inserting a TEV cleavage site between the N-terminal 6xHIS tag and the coding region. Human TFB2M was expressed as a construct lacking either only the predicted mitochondrial localization sequence (res. 1-20; Δ 20 TFB2M) or lacking the N-terminal 62 residues (Δ 62 TFB2M). The construct was generated from the pTYB11-based expression vector described previously (Sologub et al., 2009) by removing the intein tag and replacing it with a TEV-cleavable N-terminal 6xHIS tag.

Compared to wild type TFB2M, the TFB2M crystallization construct TFB2M^{cryst} lacks 62 amino acids at the N-terminus and a loop region between residues 268 and 294, for which no electron density could be observed in the IC and which was replaced by a short GSSG linker. Expression plasmids for structure determination were generated by “Around the Horn” PCR. All TFB2M variants used in transcription assays were generated using QuikChange Mutagenesis kit (Agilent) in N-his Δ 20 TFB2M background. The R601E mtRNAP variant was generated using Δ 119 mtRNAP background.

Crystallization of human TFB2M

Crystals of TFB2M could only be obtained using a truncated variant of the protein, TFB2M^{cryst} (see above for details). TFB2M^{cryst} was crystallized by the hanging drop vapour diffusion method at 20 °C by mixing 1.5 μ l of protein solution (13 mg/ml) with 1.5 μ l of reservoir solution

containing 100 mM HEPES pH 7.5, 200 mM NaCl, 17% PEG3350 and 0.4 μ l of seeding solution produced from previously grown crystals in a similar condition. Crystals were cryo-protected by gradually increasing the Glycerol concentration in the drop to 25% (v/v) final and flash frozen in liquid nitrogen.

Reconstitution and crystallization of the IC

HPLC-purified synthetic DNA oligonucleotides were obtained from IDT. Scaffolds for crystallization were annealed by heating complimentary oligonucleotides (LSP: HH NT2 / HH TS1; HSP: HH NT6 / HH TS3; Br1: HH NT2 / HH TS1-Br1; Br2: HH NT-Br2 / HH TS50; Br3: HH NT2 / HH TS-Br3; Br4: HH NT-Br4 / HH TS1; Br5: HH NT2 / HH TS-Br5; Br6: HH NT67 / HH TS-Br6) to 95°C and step-wise cooling to 4°C (1°/90s) at a final concentration of 0.5 mM in H₂O.

The human IC was reconstituted by incubating tagless Δ 104 mtRNAP (35 μ M) with a 1.1-fold molar excess of scaffold DNA, equimolar amounts of tagless TFAM (43-245 C49S) and a 1.5-fold molar excess of tagless Δ 20 TFB2M in a buffer containing 30 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM MgCl₂, 5% glycerol and 10 mM DTT for 20 min at 20 °C. The complex was subsequently purified by size exclusion chromatography using a Superdex 200 Increase 3.2/300 column equilibrated with complex crystallization buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 20 mM MgCl₂, 10 mM DTT, 1 mM TCEP). Crystals were obtained by the hanging drop vapour diffusion method at 20 °C by mixing equal volumes of protein solution and reservoir solution containing 100 mM BIS-TRIS pH 6.0, 200 mM L-Proline and 5-7% PEG8000. Crystals were cryo-protected by gradually increasing the glycerol concentration in the drop to a final of 25% (v/v) and flash-frozen in liquid nitrogen. For selenomethionine labelling and subsequent crystallization, a construct of mtRNAP lacking the N-terminal 108 amino acids (Δ 108 mtRNAP) and Δ 62 TFB2M were used.

Data collection, structure determination and refinement

Diffraction data were collected at beamline X06SA and X10SA at the Swiss Light Source in Villigen, Switzerland, with an EIGER 16M detector (Dectris) or a PILATUS 6M detector (Dectris), respectively, and at beamline P14 operated by EMBL Hamburg at the PETRA III storage ring (DESY Hamburg, Germany) with a Pilatus 6M detector (Dectris). The data was processed using XDS (Kabsch, 2010) and scaled with XSCALE (Kabsch, 2010). For selenomethionine and 5-Bromo-Uracil containing crystals, multiple datasets (from a single crystal or multiple isomorphous crystals) were merged using XSCALE to improve the anomalous signal.

The crystals of TFB2M^{cryst} belonged to space group P2₁ and diffracted to a resolution of 1.75 Å with two copies of the protein in the asymmetric unit. The structure of TFB2M^{cryst} was solved by molecular replacement in Phaser (McCoy et al., 2007) using a partial model of the yeast homolog Mtf1 (PDB ID: 1I4W, residues 134-138; 143-198; 242-281) (Schubot et al., 2001) truncated to poly-alanine. Density modification and building of an initial model was done using

phenix.autobuild (Adams et al., 2010) and subsequently completed manually in Coot (Emsley et al., 2010). The model was subjected to iterative cycles of refinement in phenix.refine (Adams et al., 2010) and manual model building in Coot until excellent stereochemistry and a free R-factor of 20.9% was obtained (Table 12). The final model contains residues 72-90, 97-267 and 295-396 of the wild-type TFB2M.

The IC crystals belonged to space group $P2_1$, contained two copies of the complex in the asymmetric unit, and showed diffraction to 4.5 Å resolution. Initial phases were obtained by molecular replacement using Phaser (McCoy et al., 2007) in the PHENIX suite (Adams et al., 2010) and the human mitochondrial transcription EC (PDB ID: 4BOC) (Schwinghammer et al., 2013) lacking nucleic acid as search model. The solution was subsequently used as starting phases for molecular replacement combined with single-wavelength anomalous diffraction (MR-SAD) and automated density modification in PHENIX using diffraction data collected from crystals containing selenomethionine-labeled mtRNAP, TFAM and TFB2M and LSP DNA. (Table 13) The resulting electron density map showed clear features of nucleic acids and proteins in addition to the search model used and was phase extended using the higher resolution native dataset in phenix.autobuild. Interpretation of the electron density was facilitated by anomalous scattering from selenium and bromine atoms incorporated into proteins and DNA, respectively (Figure 11A and B, Table 13 and Table 14). Anomalous difference Fourier maps were computed as log-likelihood gradient maps in Phaser (within the PHENIX suite) using phases derived from the refined IC model (Read and McCoy, 2011). Modelling of the IC was done largely using the experimental map and cross-validated with a map generated by MR-SAD with the keyword “phaser_sites_then_phase=True” to obtain a map free of model bias.

Individual domains and secondary structure elements of mtRNAP were rigid body fitted in real space manually to fit the experimental electron density. The region of mtRNAP corresponding to the specificity loop (residues 1086-1107) showed fragmented density and was modeled based on the structure of the T7 RNAP initiation complex. (Cheetham et al., 1999) The density allowed for modelling the main chain trace of this element lacking only three residues (1094-1096) at the tip, yet the sequence register could not be assigned unambiguously and it was therefore modeled as poly-alanine.

To obtain an initial model of the IC, the crystal structure of TFAM (PDB ID: 3TMM) (Ngo et al., 2011) in complex with LSP DNA was fitted into the electron density and the individual domains were rigid body fitted locally in real space using Coot (Emsley et al., 2010). The DNA emerging from TFAM was extended and adjusted as ideal B-DNA to the expected melting point at the beginning of the mismatched region (-4) and rigid body fit locally in real space. The experimental density allowed for modelling of two additional bases of the template and three of the non-template strand, respectively, past the melting point. The downstream duplex DNA showed weaker density and was positioned with the help of anomalous difference peaks from crystals containing DNA labelled with 5-bromo-uracile at specific sites (Figure 11B). In the crystal form observed, the downstream DNA mediates a crystal contact and may therefore be stabilized in the observed conformation.

To model TFB2M, the crystal structure of TFB2M^{cryst} was rigid body fit into the experimental IC density. Correct positioning of TFB2M was verified using peaks in an anomalous difference Fourier map calculated from the dataset which was used for phasing the IC, which included selenomethionine-labelled TFB2M (Figure 11A). Placement of TFB2M led to a single clash between the C-terminal tail of TFB2M with the mtRNAP intercalating hairpin and the non-template DNA (Figure 17). Since this C-terminal region appears to be flexible, the TFB2M model was truncated to the last residue with clear density in the IC map (residue 392).

After positioning of all known protein structures, a residual unexplained density in the experimental map remained close to the HMG box B of TFAM with three weak peaks in the anomalous difference map for selenium. An anomalous difference map calculated from a dataset obtained from crystals in which only mtRNAP was selenomethionine-labeled indicated that these peaks originate from residues within the polymerase. Based on a unique “MRM” sequence motif found in the thus far not observed N-terminal extension region of mtRNAP and secondary structure prediction using PSIPRED (Buchan et al., 2013), a helix spanning residues 122-146 in mtRNAP was assigned to the unmodeled density, which we termed “tether helix” (Figure 15A).

Refinement of the IC model against the native dataset using phenix.refine with secondary structure restraints, reference model restraints and DNA geometry restraints resulted in a model with good geometry and a free R-factor of 30.9%. (Table 13) The resulting mFo-DFc map showed difference density for some additional features such as the polypeptide path connecting the mtRNAP tether helix to the PPR domain and parts of the missing single-stranded non-template DNA strand, but we refrained from modelling these features due to the limited resolution of the data.

The structure of the HSP IC was solved by molecular replacement using the LSP IC as search model and subsequently adjusted to the HSP sequence. The model was refined in phenix.refine using a similar protocol as for the LSP IC and led to a final model with good stereochemistry and a free R of 33.1%. (Table 15) The identical orientation of TFAM on both LSP and HSP was verified by comparing anomalous difference peaks calculated using structure factors from crystals containing selenomethionine-labeled TFAM and the respective promoter DNA. (Figure 15B).

Figures were prepared using PyMol. Surface charge analysis was performed using the APBS plugin for PyMol (Baker et al., 2001) and displayed with +/- 1kT/e.

2.4. Mechanism of transcription anti-termination in human mitochondria

This section contains methods specifically related to the study on the mechanism of transcription anti-termination in human mitochondria. Only those methods developed and performed as part of this thesis are described here (unless stated otherwise), while methods carried out predominantly by the collaborators from the Temiakov lab are described in section 5.2.1. Parts of this section have been published:

H.S. Hillen, A.V. Parshin, K. Agaronyan, Y.I. Morozov, J.J. Graber, A. Chernev, K. Schwinghammer, H. Urlaub, M. Anikin, P. Cramer and D. Temiakov (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell*, in press

A detailed list of author contributions can be found on page VI.

Expression and purification of the components of human mitochondrial transcription

Human mtRNAP ($\Delta 119$ and $\Delta 150$), WT TFAM, $\Delta 20$ TFB2M, WT TEFM, and $\Delta 50$ TEFM (all with non-cleavable 6xHIS-tag) were expressed and purified as described previously (Agaronyan et al., 2015; Sologub et al., 2009) with the adaptations described in 2.2.2. Variants of these proteins were obtained by site-directed mutagenesis (QuikChange, Agilent) and purified as described previously (Morozov et al., 2015). TEFM NTD (residues 36-136) was purified by the Temiakov lab using cation exchange chromatography on a MonoS column followed by gel-filtration on a Superdex 200 10/300 column. Expression of Bpa-containing proteins was performed by the Temiakov lab as described previously (Morozov et al., 2015). Selenomethionine-labeled TEFM was purified as described above.

Crystallization of human TEFM C-terminal domain

Purified $\Delta 50$ TEFM (6.5 mg/ml) was treated with ArgC protease (Promega, 1:1000 w/w) and incubated for 60 min at room temperature prior to crystallization trials. Initial hits were obtained using the sitting drop vapor diffusion with a well solution containing 100 mM BIS-TRIS pH 5.5 – 6.5, 200 mM MgCl₂ and 25% PEG3350. Large, diffraction-quality crystals were obtained by the hanging drop vapor diffusion method by micro-seeding with crushed crystals into fresh drops. Crystals appeared within 24 h and grew to full size over the course of two weeks. Crystals were cryo-protected by gradually increasing the glycerol concentration to 20% (v/v) and flash frozen in liquid nitrogen.

Crystallization of the EC/TEFM complex

Synthetic DNA and RNA oligonucleotides (IDT DNA) HH NT25, HH TS22 and HH RNA14 were dissolved in H₂O and mixed at a final concentration of 0.5 mM. The scaffold was annealed by heating the mixture to 95°C and step-wise cooling (1 °C / 90 s) to 4 °C. To form an EC/TEFM complex, Δ 150 mtRNAP (45 μ M) was initially incubated with a 1.1- fold molar excess of DNA/RNA scaffold and incubated at 20 °C for 10 min prior to addition of a 2.2-fold molar excess Δ 135 TEFM and further incubation at 20°C for 10 min. The complex was subsequently purified by size exclusion chromatography using a Superdex 200 Increase 3.2/300 column (GE Healthcare) equilibrated with a buffer containing 10 mM Tris-HCl pH 8.0 at 23 °C, 100 mM NaCl, 20 mM MgCl₂, 10 mM DTT, 1 mM TCEP and 5% (v/v) Glycerol. Peak fractions were pooled and used for crystallization. Crystals were obtained by the hanging-drop vapour diffusion method by mixing 2.5 μ l of protein complex solution with 2.5 μ l of reservoir solution containing 100 mM BIS-TRIS pH 5.5, 200 mM (NH₄)₂SO₄, 6% PEG3350 and 0.5 μ l of a seeding solution prepared from crystals previously obtained in an identical condition. Crystals appeared as thin blade clusters after 3-4 days. The clusters were carefully broken into individual blades and cryoprotected by gradually increasing the glycerol concentration in the drop to 25% (v/v) final prior to fishing and freezing in liquid nitrogen.

Structure determination, model building and refinement

Diffraction data for TEFM CTD and EC-TEFM complex crystals was collected at beamlines X06SA and X10SA at the Swiss Light Source (Villigen, CH) using a Pilatus 6M detector. Diffraction data for TEFM NTD crystals was collected at beamline P13 at the PETRA III storage ring operated by EMBL (DESY, Hamburg, Germany). Data were indexed and integrated using XDS and scaled using XSCALE (Kabsch, 2010). Anomalous substructure determination, phasing and density modification was performed using the hkl2map interface and SHELXC/D/E (Pape and Schneider, 2004; SheldrickIUCr, 2008), as well as phenix.autosol (Adams et al., 2010). To obtain experimental phase information for the TEFM CTD crystals, multiple datasets from two isomorphous selenomethionine-containing crystals were merged using XSCALE and used for SAD phasing. Heavy atom sites identified by SHELXD were manually edited to remove weak occupancy sites and used as input in phenix.autosol for phasing and density modification. The obtained electron density map was of excellent quality and allowed for manual building of most of the four protein chains in the asymmetric unit. (Figure 21) Refinement was carried out against the high-resolution native dataset using phenix.refine (Adams et al., 2010) and alternated with manual adjustment in Coot (Emsley et al., 2010) until a model with excellent stereochemistry was obtained. Experimental phase information for TEFM NTD crystals was obtained in a SAD experiment utilizing native sulfur atoms present in cysteine and methionine residues by collecting multiple highly redundant datasets from a single crystal at a wavelength of 2.066 Å. Datasets were merged using XSCALE and substructure determination, phasing, density modification and phase extension was carried out with SHELXC/D/E. The resulting electron density map was readily interpretable and used for manual model building. The model was subjected to iterative rounds of refinement in

phenix.refine against the high-resolution native dataset collected from the same crystal and manual model building in Coot until excellent stereochemistry and R-factors were achieved. The final mFo-DFc map displayed some unmodeled difference density between crystallographically related molecules, corresponding to unidentified solvent molecules.

The EC-TEFM structure was solved by molecular replacement in Phaser (McCoy et al., 2007) using the human mitochondrial transcription elongation complex (PDB ID: 4BOC) (Schwinghammer et al., 2013) and the TEFM CTD structure as search models. The resulting electron density was readily interpretable and used to rigid body fit domain movements in mtRNAP manually in Coot. The model was subsequently refined in phenix.refine alternated with manual adjustment in Coot, with secondary structure restraints and reference model restraints enabled until the late stages of refinement due to the low resolution of the data. We observed positive difference density corresponding to the previously unresolved single-stranded region of the non-template strand as well as for parts of the N-terminal extension of mtRNAP, which may be involved in forming crystal contacts, but refrained from *de novo* modeling these regions due to the limited resolution. The unbiased Fo-Fc map for the non-template strand was calculated from a model which never contained any model for the non-template strand. The final model was refined to good stereochemistry and a free R factor of 27.6%.

Mapping of TEFM-mtRNAP interaction using BS3 and EDC cross-linking

The EC/TEFM complex was assembled by incubating 10-20 μ M Δ 150 mtRNAP with a 2.5-fold molar excess of WT TEFM and a 1.3-fold molar excess of HH RNA14/HH NT19/HH TS16 scaffold for 10 min at 20°C. The complex was subsequently purified by size exclusion chromatography using a Superdex 200 Increase 3.2/300 column (GE Healthcare) equilibrated with buffer containing 20 mM HEPES (pH 8.0 for BS3 or pH 7.5 for EDC), 100 mM NaCl, 20 mM MgCl₂, and 0.5 mM TCEP. The purified complexes (~2 μ M) were then incubated with a 150 to 470-fold molar excess of BS3 (Thermo Fisher) for 30 min at 30 °C or with 2500 to 15000-fold molar excess of EDC (Thermo Fisher) in the presence of 1 mM N-hydroxy-sulfosuccinimide for 60 min at 23°C. The reactions were quenched by the addition of 200 mM (NH₄)₂HCO₃ followed by incubation for 5 min on ice. The products of the reaction were resolved using 4-12% SDS-PAGE.

The following steps were subsequently performed by the Urlaub lab: the cross-linked species excised and in-gel digested with trypsin. The tryptic peptides were extracted, dried and reconstituted in solution containing 5% v/v acetonitril and 0.1% v/v FA and subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS). The separated peptides were analyzed online on a Q-Exactive HF mass spectrometer (Thermo Scientific). The raw data of LC-MS/MS analysis were converted to Mascot generic format files with Proteome Discoverer 2.0.0.802 software (Thermo Scientific) and searched against a protein database by pLink 1.22 software (Yang et al., 2012) as described in (Pleiner et al., 2015). The results were obtained

with 1% false discovery rate followed by manual validation and removal of all tandem spectra that did not show confident sequence coverage of both cross-linked peptides.

Analytical size-exclusion chromatography

The mitochondrial initiation complex was assembled by incubating Δ 119 mtRNAP (20 μ M) with the LSP promoter template (20 μ M) (template strand: HH TS1; non-template strand: HH NT2), Δ 42 TFAM (43-256, C49S) (20 μ M) and Δ 20 TFB2M (40 μ M) for 10 min at 20°C. To analyze binding of TEFM, WT TEFM (60 μ M) was added to the IC and incubated for 10 min at 20°C. Gel-filtration was performed using a Superdex 200 Increase 3.2/300 size exclusion column (GE Healthcare) equilibrated with buffer containing 20 mM Tris-HCl pH 8.0 at 23 °C, 100 mM NaCl, 20 mM MgCl₂, and 2 mM DTT. For the dimer mutant analysis, the ECs (20 μ M) were assembled using R14/TS22HH/NT25HH scaffold (1.1-fold molar excess) for 10 min at 20°C. After incubation, the ECs were mixed with 2.2-fold molar excess of WT TEFM or TEFM mutant (F244E L248D I252D M256S L260D) and incubated for 10 min at 20°C. Gel-filtration was performed using a Superdex 200 3.2/300 size exclusion column (GE Healthcare) equilibrated with buffer containing 20 mM Tris-HCl pH 8.0 at 23 °C, 200 mM NaCl, 20 mM MgCl₂, 10 mM DTT and 1 mM TCEP.

3. Results and Discussion

3.1. Structural basis of mitochondrial transcription initiation

Results presented in this section were obtained in collaboration with the Temiakov lab at the Department of Cell Biology, Rowan University, Stratford, NJ, USA and are currently under peer review for publication:

H.S. Hillen, Y.I. Morozov, A. Sarfallah, D. Temiakov and P. Cramer (2017) Structural basis of mitochondrial transcription initiation. *Cell*, in revision

A detailed list of author contributions can be found on page VI.

For clarity, this results section presents all results of this publication. The structural data was obtained as part of this thesis work, while the biochemical data was generated by the collaborators. Methods performed as part of this thesis are described in section 2.3 and methods predominantly performed by the collaborators are presented in section 5.1.1.

3.1.1. Abstract

Transcription in human mitochondria is driven by a single-subunit, factor-dependent RNA polymerase (mtRNAP). Despite of its critical role in both expression and replication of the mitochondrial genome, transcription initiation by mtRNAP remains poorly understood. Here we report crystal structures of human mitochondrial transcription initiation complexes assembled on both light and heavy strand promoters. The structures reveal how transcription factors TFAM and TFB2M assist mtRNAP to achieve promoter-dependent initiation. TFAM tethers the N-terminal region of mtRNAP to recruit the polymerase to the promoter, whereas TFB2M induces structural changes in mtRNAP to enable promoter opening and trapping of the DNA non-template strand. Structural comparisons demonstrate that the initiation mechanism in mitochondria is distinct from that in the well-studied nuclear, bacterial, or bacteriophage transcription systems, but that similarities are found on the topological and conceptual level. These results provide a framework for studying the regulation of gene expression and DNA replication in mitochondria.

3.1.2. Results

Structure of human TFB2M

To investigate the mechanism of mitochondrial transcription initiation, we first set out to complete the set of structures for proteins involved in mitochondrial transcription initiation and determined the structure of TFB2M. As extensive crystallization trials using full-length human TFB2M did not yield crystals, we designed a variant lacking putatively flexible regions that may impair crystallization. This variant, TFB2M^{cryst}, lacks 62 N-terminal residues and a predicted internal loop (residues 268-294) that was replaced by a short GSSG-linker. Functional characterization of this TFB2M variant shows that replacement of the internal loop does not affect the transcriptional activity of TFB2M (Figure 7A), whereas the N-terminal truncation is known to reduce the activity of TFB2M due to its role in interactions with the priming nucleotide (Figure 7A) (Sologub et al., 2009). TFB2M^{cryst} yielded crystals that diffracted to 1.75 Å resolution (Figure 7B and C). The structure was solved by molecular replacement (Table 12). The final model shows very good stereochemistry and contains residues 72-396 of TFB2M with the exception of a short linker (91-96).

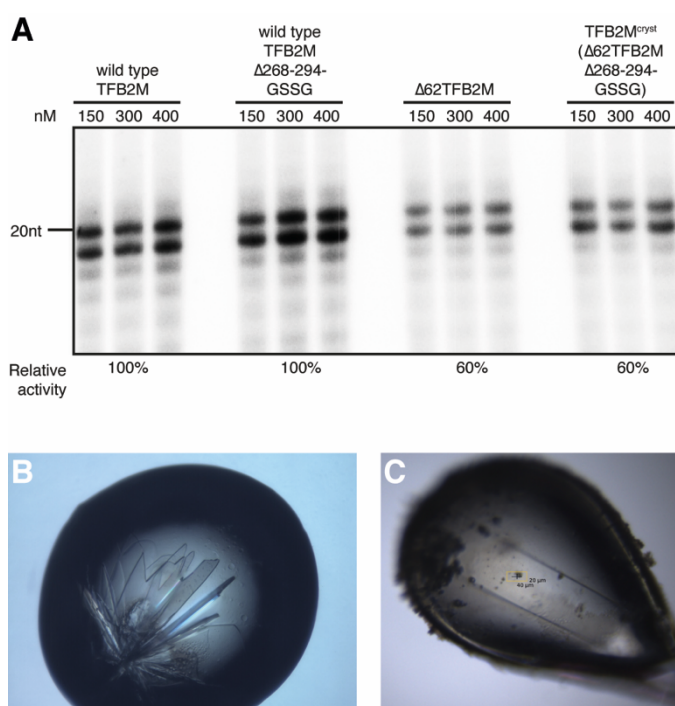


Figure 7: Biochemical characterization and crystallization of a human TFB2M variant.

(A) Functional characterization of the effect of truncations used to obtain a crystallizable TFB2M variant. Replacement of the loop between $\beta 6$ and $\beta 7$ (residues 268-294) does not affect transcriptional activity, while removal of 62 amino acids from the N-terminus results in a decrease of activity by 40%, consistent with previous reports that the N-terminal region of TFB2M plays a role in priming RNA synthesis (Sologub et al., 2009). (B) Exemplary initial crystal hit from commercial crystallization screens in a condition similar to the final crystal growth condition. (C) Optimized crystal used for structure determination.

The structure (Figure 8) shows that TFB2M resembles the paralogous mitochondrial methyltransferase TFB1M (Guja et al., 2013) and the yeast mitochondrial transcription

initiation factor Mtf1 (Schubot et al., 2001) (Figure 9A and B). As predicted from sequence homology, the N-terminal domain (residues 72-305) adopts a fold resembling S-adenosyl-methionine-dependent methyltransferases with a central seven-stranded β -sheet flanked on either side by three α -helices (Martin and McMillan, 2002). Similar to TFB1M and Mtf1, TFB2M deviates from the canonical methyltransferase fold by an insertion between $\beta 6$ and $\beta 7$, which corresponds to the region replaced with the GSSG linker in the crystallization construct (Guja et al., 2013; Schubot et al., 2001). In addition, TFB2M displays a prominent loop insertion between $\beta 3$ and $\alpha 4$ not found in either of the two other proteins. The C-terminal domain (residues 306-396) consists of four alpha helices and an extended C-terminal tail (residues 389-396), which is likely flexible in solution as clear density for this region was only observed for one of the two copies in asymmetric unit. The structure of TFB2M completes the set of high-resolution structures of proteins involved in mitochondrial transcription initiation.

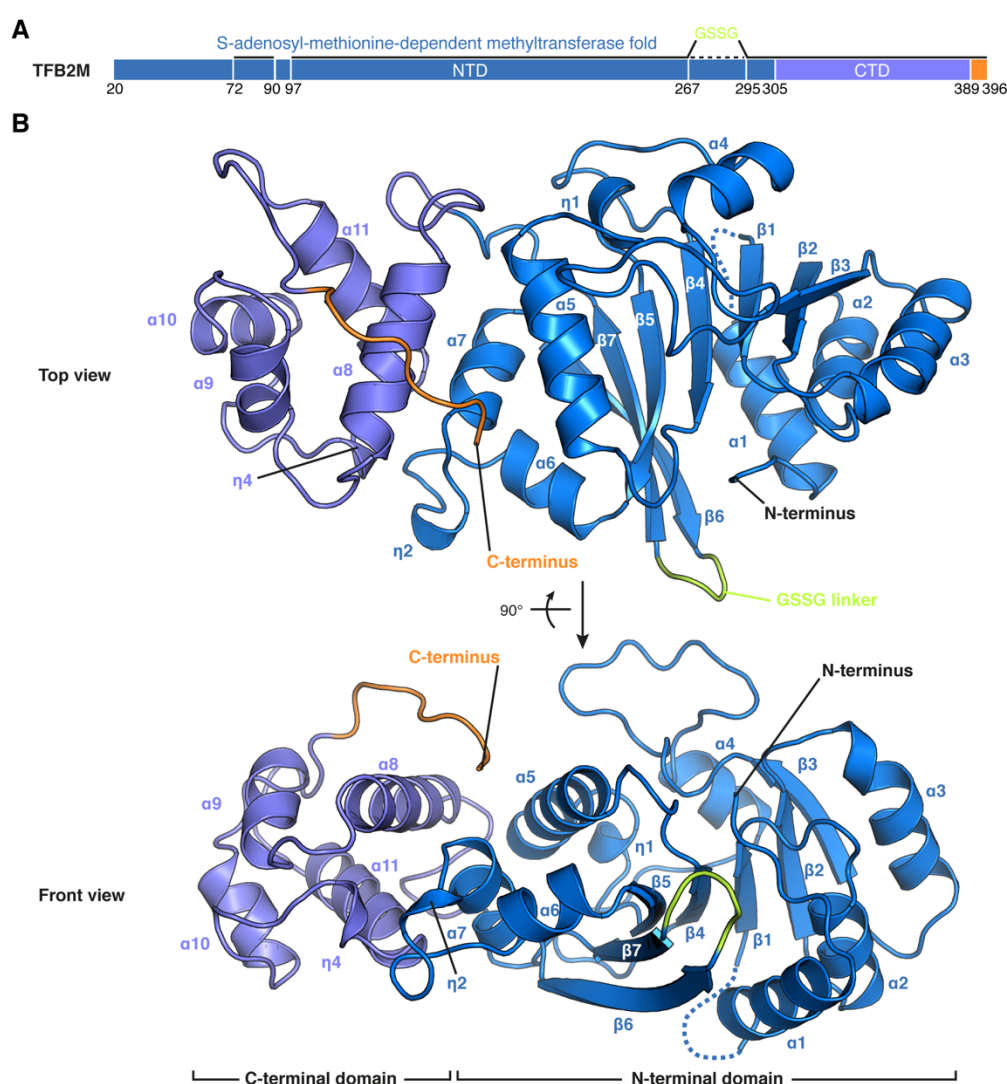


Figure 8. Structure of human TFB2M.

(A) Schematic representation of TFB2M. The N-terminal domain is colored in marine blue, the C-terminal domain in slate and the C-terminal tail in orange. Regions visible in the electron density of the TFB2M^{cryst} crystals are indicated by a solid black line. The dashed black line represents the internal loop region replaced by a GSSG-linker in order to obtain a crystallizable construct. (B) Ribbon representation of the human TFB2M structure. Coloring as in (A) with secondary structure elements indicated.

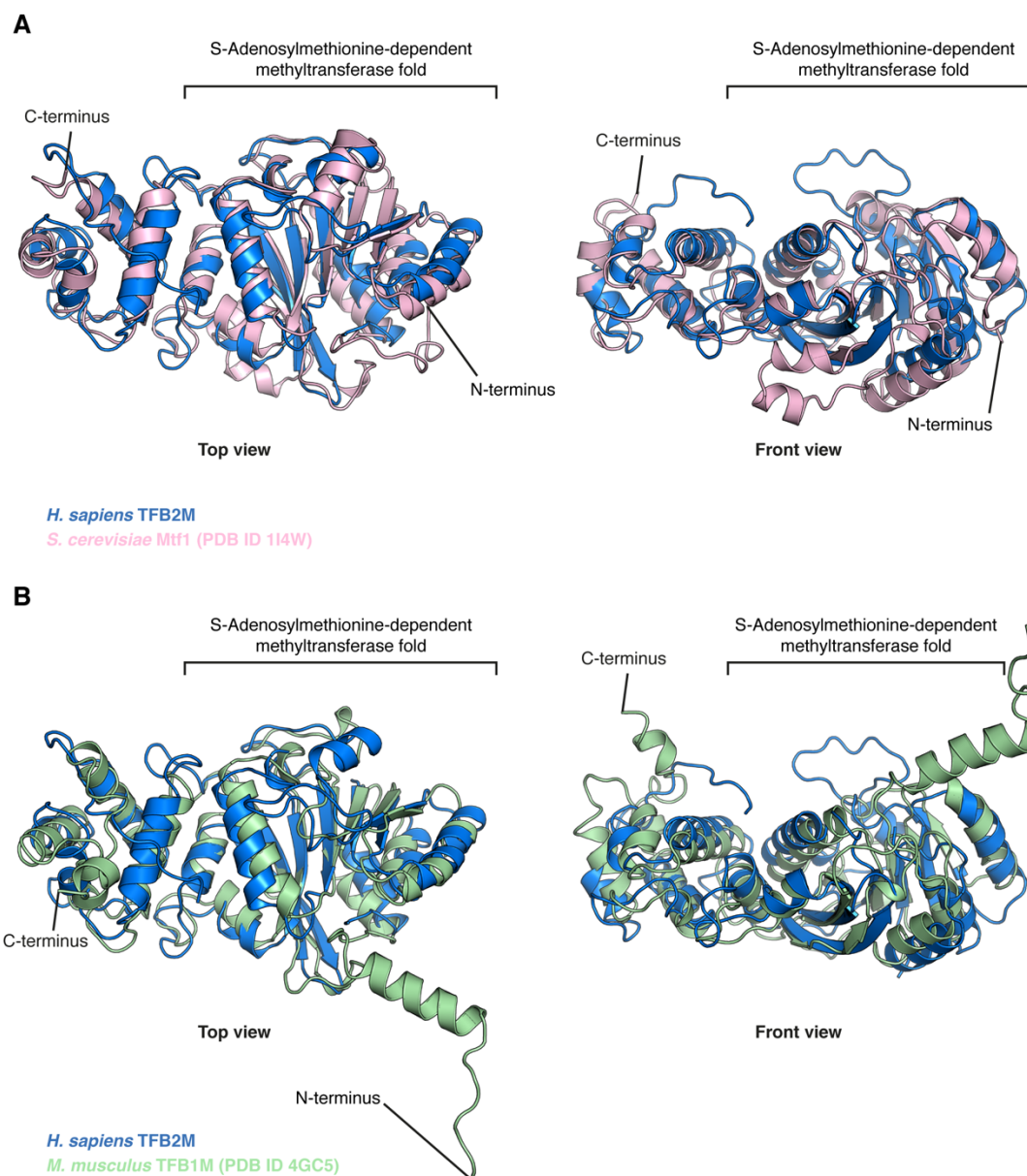


Figure 9. Structural comparison of TFB2M homologs.

(A) Structural comparison of human TFB2M (marine blue) and *S.cerevisiae* Mtf1(lightpink) (PDB ID: 1I4W) (Schubot et al., 2001). The proteins were aligned over 256 residues with an r.m.s.d of 3.9 Å. While both proteins have a loop insertion between $\beta 6$ and $\beta 7$, human TFB2M has an additional loop insertion between $\beta 3$ and $\alpha 4$. Note that the C-terminal tail is not fully resolved in the structure of Mtf1.

(B) Structural comparison of human TFB2M (marine blue) and *M.musculus* TFB1M (palegreen) (PDB ID: 4GC5) (Guja et al., 2013). The proteins were aligned over 264 residues with an r.m.s.d of 3.9 Å. Compared to TFB2M and Mtf1, the insertion between $\beta 6$ and $\beta 7$ is small in TFB1M. Furthermore, TFB1M displays a protruding alpha-helix in the N-terminal region which is not predicted to be present in TFB2M. Note that the C-terminal tail is not fully resolved in the structure of TFB1M and the N-terminal region is not fully resolved in TFB2M.

Structure determination of the mitochondrial transcription initiation complex

We then assembled a transcriptionally active IC consisting of TFAM, TFB2M, mtRNAP, and either LSP or HSP DNA containing a pre-melted region spanning register -4 to +3, which corresponds to the DNA region initially unwound around the transcription start site +1 (Ramachandran et al., 2016) (Figure 10A and B). After extensive optimization, crystals of the IC were obtained that diffracted to 4.5 Å resolution (Figure 10C and D). The IC crystal structure was determined by a combination of molecular replacement and anomalous diffraction (Methods and Table 13, Table 14), which led to an interpretable electron density map (Figure 11C). The known structures of mtRNAP (Schwinghammer et al., 2013) and TFAM (Ngo et al., 2011) were fitted into the electron density and the newly obtained TFB2M structure could be unambiguously placed. Correct positioning of TFB2M was verified using an anomalous difference Fourier map that revealed selenium peaks for all nine methionine residues (Figure 11A). Most of the DNA could be built, except for parts of the single-stranded region, and the correct sequence register was confirmed using anomalous diffraction from 5-Bromo-Uracil labeled DNA scaffolds (Figure 11B and Table 14). This led to an atomic model for the IC refined to a free R-factor of 30.9% (Table 13). We also solved a 4.5 Å resolution crystal structure of the IC assembled on the HSP promoter (Table 15). This structure was essentially identical to the LSP IC (r.m.s.d. = 0.4 Å over 10,531 atoms) (Figure 11D) and in the following we discuss only the LSP IC.

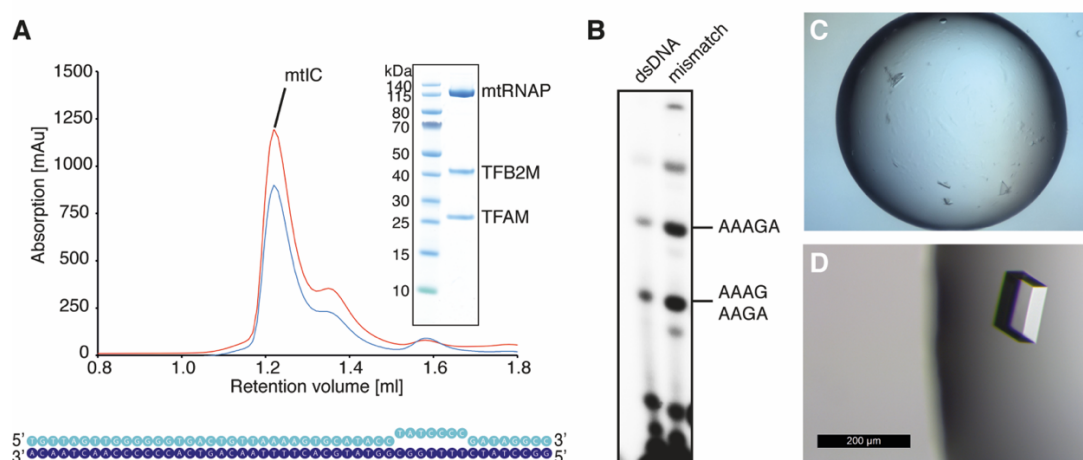


Figure 10. Reconstitution and crystallization of the human mitochondrial IC.

(A) Reconstitution of the human mitochondrial transcription initiation complex. Shown is a representative size exclusion chromatogram of the final purification step prior to crystallization. The peak containing the IC is indicated. SDS-PAGE analysis of the peak fraction demonstrates stoichiometric presence of mtRNAP, TFAM and TFB2M. The nucleic acid scaffold which was used to assemble the LSP IC is depicted schematically below with each base represented as a circle and the mismatch bubble around the transcription start site indicated. The template strand is colored blue, the non-template strand cyan. (B) Activity of the IC used for structure determination. Transcription initiation activity was assayed using radiolabeled ATP and a DNA template resulting in synthesis of 4- and 5-nt RNA. dsDNA: double stranded DNA; mismatch: -4 to +3 mismatched template as depicted in (A). (C) Initial crystal hit from commercial crystallization screen. (D) Exemplary image of the optimized crystal form which showed X-ray diffraction to 4.5 Å resolution.

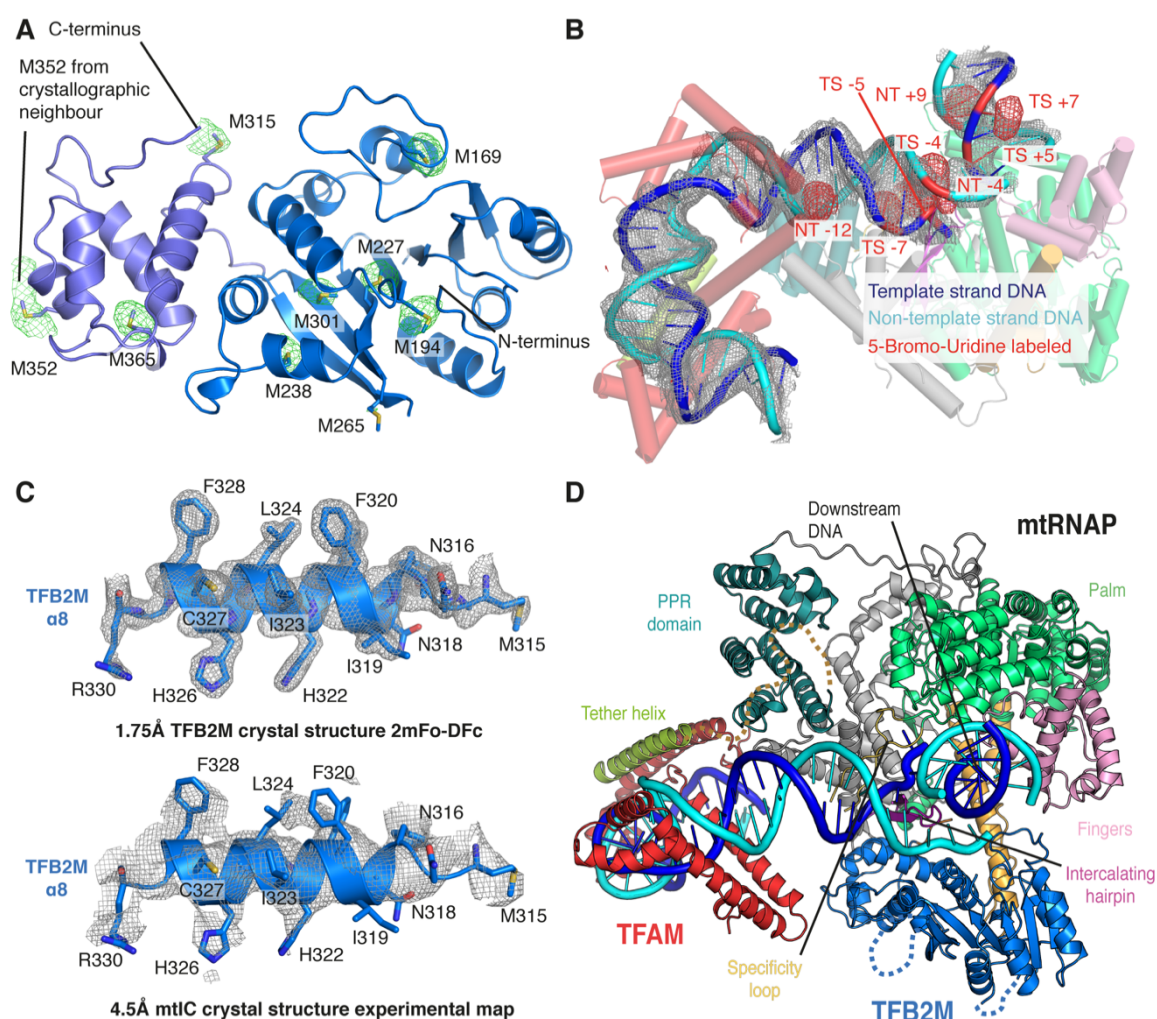


Figure 11. Structure determination of the human mitochondrial IC

(A) Positioning of the high-resolution TFB2M crystal structure in the IC structure. The TFB2M structure is shown as ribbon representation with coloring as in Figure 8. Peaks in the anomalous difference Fourier map calculated from the crystals containing selenomethionine labeled proteins are shown as green mesh at 3.5σ . All selenium peak positions coincide with methionine residues in the positioned TFB2M structure. (B) Confirmation of sequence register assignment for DNA by anomalous diffraction from 5-Bromo-Uracil labeled DNA. The final refined model of the IC is shown transparently as ribbon representation with a $2mF_o-DF_c$ map around the DNA shown as grey mesh at 1.0σ . Helices are depicted as cylinders. Anomalous difference Fourier maps are depicted as red mesh at 4.0σ . Shown are the maps from six different 5-Bromo-Uracil labeled DNA scaffolds (Table 14 and Methods). (C) Comparison of electron densities for the functionally important $\alpha 8$ helix of TFB2M. The $2mF_o-DF_c$ map for the high-resolution TFB2M structure is shown at 1.0σ (top) and the experimental electron density for the IC is shown at 1.0σ (bottom). (D) Structure of the IC assembled using HSP promoter DNA. Important structural elements are indicated and colored as for the LSP IC in Figure 12. The structure of the IC is nearly identical on LSP and HSP promoter DNA.

IC structure reveals locations of TFAM, TFB2M and DNA on mtRNAP

The IC structure (Figure 12) reveals that mtRNAP is largely unchanged compared to the previously reported EC structure (Schwinghammer et al., 2013), with the exception of the fingers domain, which adopts the “clenched” conformation observed in the apo enzyme (Ringel et al., 2011). Whereas the position of the downstream DNA duplex in the IC is identical to that observed in the EC, the upstream DNA occupies a different location, running along the NTD of mtRNAP. The conserved intercalating hairpin of mtRNAP separates the DNA strands at the upstream edge of the open DNA region observed in the active center cleft of the polymerase (Figure 12B and Figure 13A).

TFB2M contacts the intercalating hairpin and covers the junction between the upstream DNA duplex and the open DNA region (Figure 12B and Figure 13A). TFAM binds DNA 16-39 nt upstream of the transcription start site and induces a $\sim 180^\circ$ bend into DNA, resembling the free TFAM-DNA complex (Ngo et al., 2011; Rubio-Cosials et al., 2011). In agreement with cross-linking data (Morozov et al., 2014), TFAM does not contact TFB2M, but binds the N-terminal domain (NTD) of mtRNAP at helix D. In addition to the severe upstream bending of the DNA induced by TFAM binding, the trajectory of the DNA is changed by $\sim 45^\circ$ between mtRNAP and TFAM (Figure 13B), and the downstream DNA duplex encloses an angle of $\sim 135^\circ$ relative to the upstream duplex at the point of DNA melting (Figure 13B).

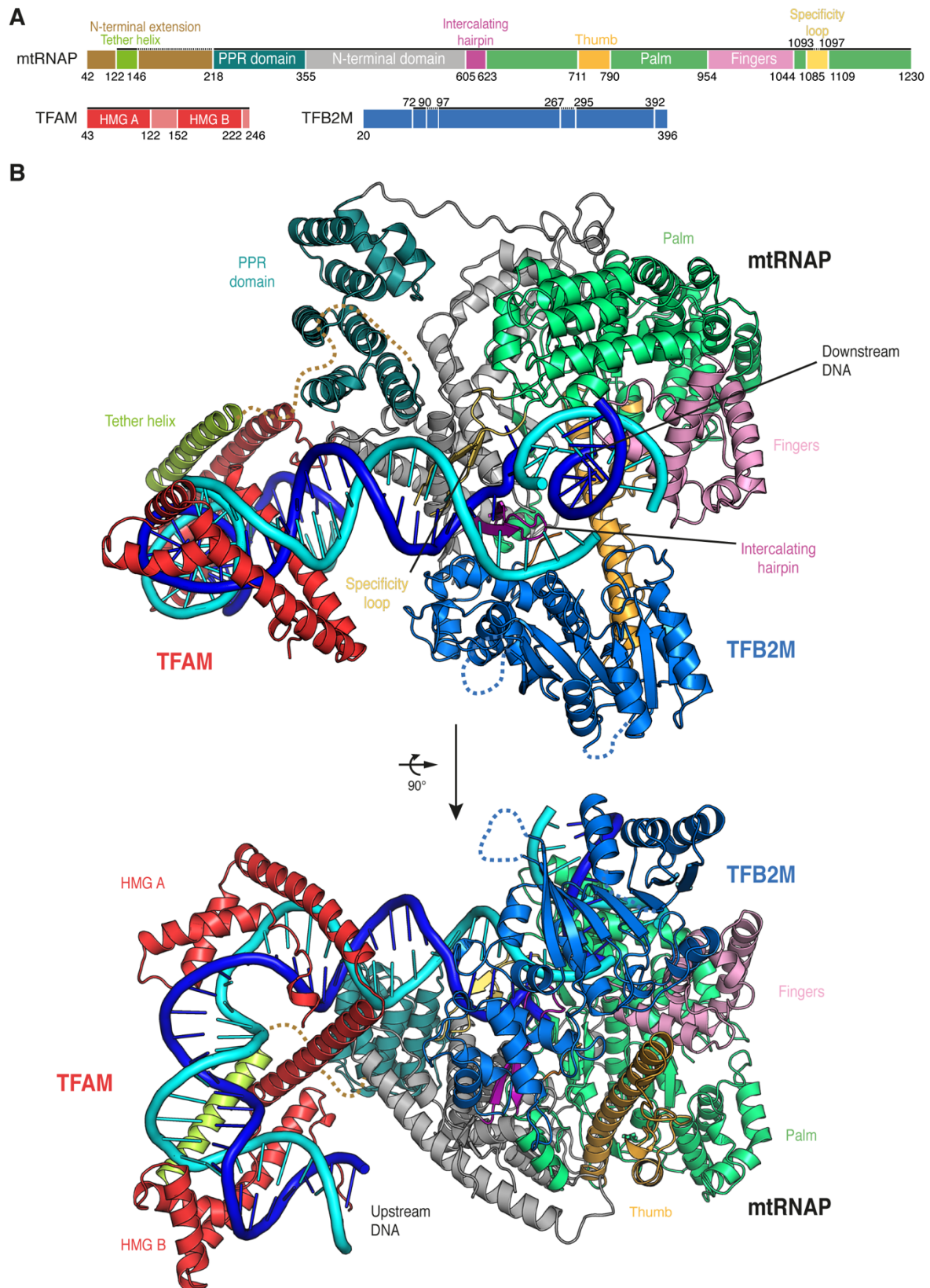


Figure 12. Structure of the human mitochondrial transcription initiation complex.

(A) Schematic representation of mtRNAP, TFAM and TFB2M. Important structural elements are indicated with flanking residue numbers. Regions with interpretable electron density in the IC crystal structure are indicated by a solid black line. Regions with density of insufficient quality for model building are indicated by a dashed black line. The color code is used throughout. (B) Ribbon representation of the IC structure assembled on LSP DNA. Important structural elements are indicated.

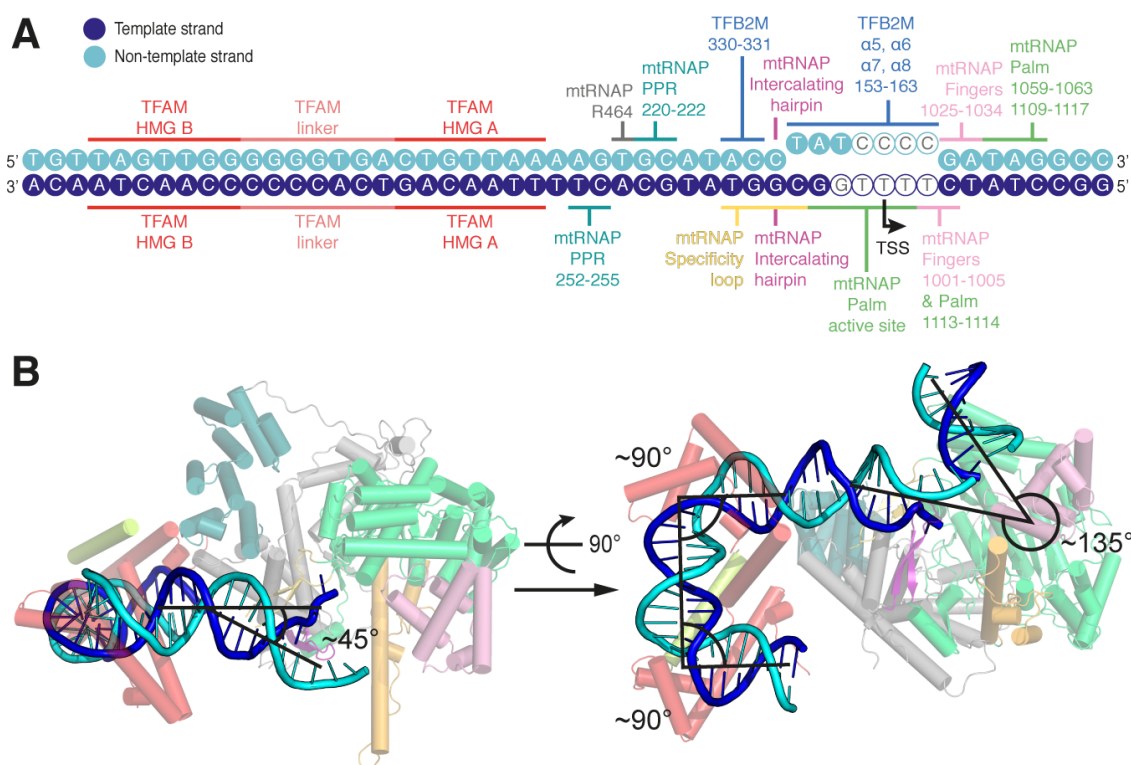


Figure 13. Promoter DNA interactions in the IC and DNA bending.

(A) Schematic representation of protein-DNA interactions. DNA bases of the LSP promoter are depicted as circles in blue and cyan for the template and non-template strand, respectively. DNA bases lacking density in the IC crystal structure are depicted as hollow circles. Protein regions interacting with the DNA are indicated with coloring as in Figure 12. (B) DNA bending observed in the IC structure. Ribbon representation of the IC LSP structure with mtRNAP and TFAM in the background and DNA in the foreground. Helices are depicted as cylinders. TFB2M was omitted for clarity. Approximate angles between DNA duplexes are indicated.

TFAM recruits mtRNAP to promoter DNA

The IC structure explains how TFAM recruits mtRNAP to promoter DNA (Gaspari et al., 2004; Morozov et al., 2014; Posse et al., 2014). The HMG Box B domain of TFAM interacts with a newly observed “tether” helix in the N-terminal extension of mtRNAP, thereby anchoring mtRNAP to the promoter (Figure 14 and Figure 15A). The C-terminal tail of TFAM is located close to the PPR domain and residues 444-462 of mtRNAP (D-helix), consistent with published biochemical, genetic and cross-linking data (Dairaghi et al., 1995b; Morozov and Temiakov, 2016; Morozov et al., 2015) (Figure 14). These contacts enable TFAM to recruit mtRNAP and position its active site over the transcription start site for *de novo* RNA synthesis (Dairaghi et al., 1995a; Morozov et al., 2014). In agreement with cross-linking data (Morozov and Temiakov, 2016), TFAM binding is identical in the structure of the HSP IC (Figure 15B). There, similarly to the LSP IC, TFAM binds to the region that is located 16-39 bp upstream to the HSP transcription start site, in agreement with earlier foot-printing data (Fisher et al., 1987). This indicates that the two transcription units in human mitochondria possess similar architecture, in contrast to previous reports that suggested no role of the TFAM C-terminal tail in LSP activation (Uchida et al., 2017) and opposite orientations of TFAM relative to mtRNAP in the IC assembled on HSP DNA (Ngo et al., 2014).

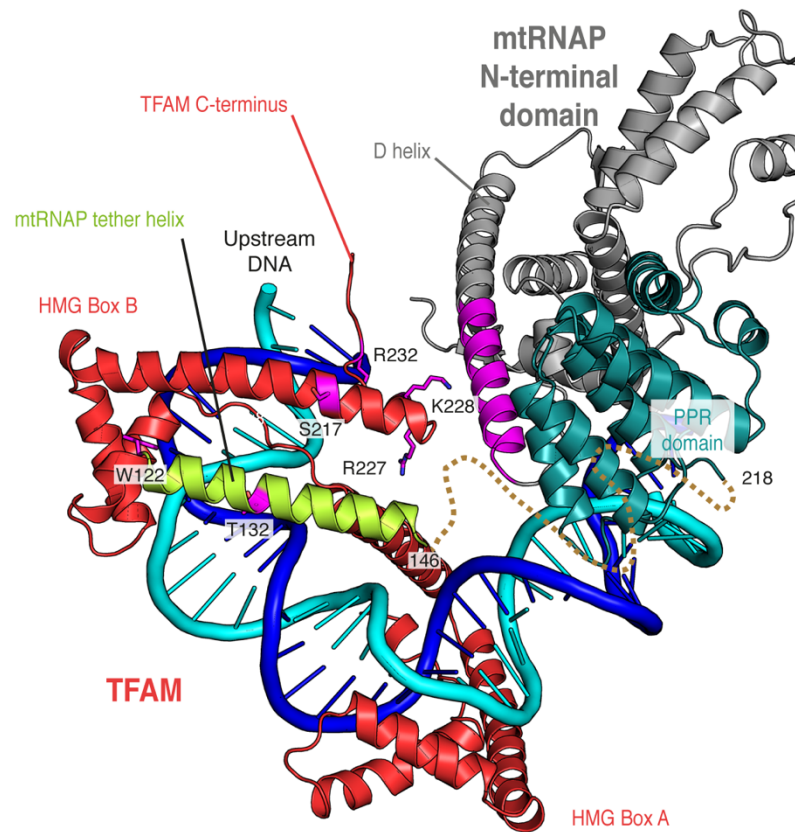


Figure 14. TFAM recruits mtRNAP to promoter DNA.

Close-up view of the interaction between mtRNAP and TFAM in the LSP IC. Coloring as in Figure 12. Residues in TFAM and the region in mtRNAP previously shown to be functionally important for initiation or identified as crosslinking-points are shown in magenta (Morozov et al., 2014; 2015). The trajectory of the polypeptide chain of the N-terminal extension based on observed difference density is indicated as a dashed line.

that are essential for activity (Morozov et al., 2015), including residue H326, which is critical for transcription initiation (Figure 16A and C and Figure 17).

The PBD also harbors a “lever” loop (residues 588-604), a structural element which is adjacent to the intercalating hairpin and found in mtRNAP but not in phage RNAPs. The lever loop is essential for initiation (Morozov et al., 2015) and likely plays a key role in TFB2M-induced rotation of the core NTD. The lever loop would clash with bound TFB2M if it adopted the position observed in free mtRNAP (Figure 16A). In the IC, the lever loop interacts with loop $\alpha 9$ - $\alpha 10$ in TFB2M (residues 341-347), and this may stabilize the rotated NTD core. Indeed, mutation of an arginine residue in the lever loop (R601E) results in decreased transcription initiation (Figure 16C).

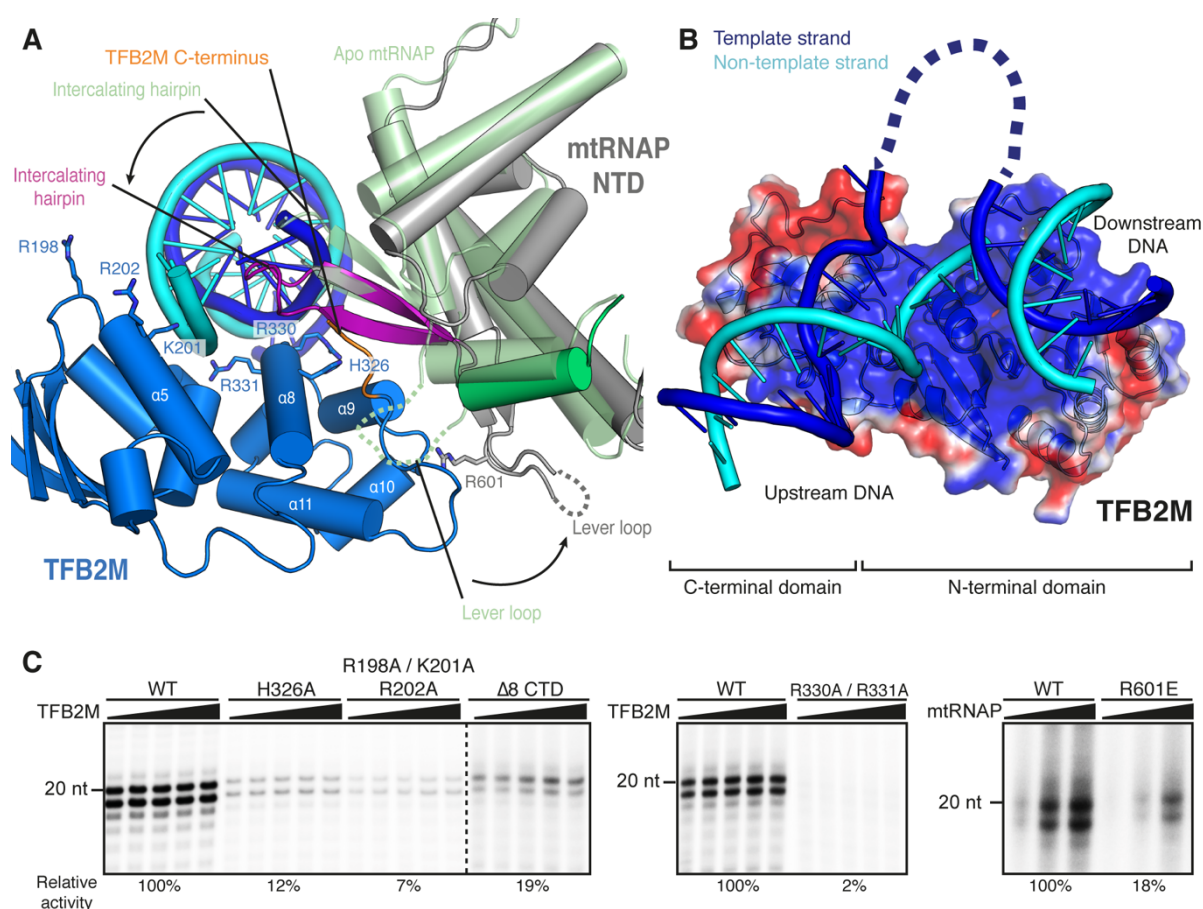


Figure 16. TFB2M stabilizes open DNA and traps the non-template strand.

(A) TFB2M binding induces structural rearrangements in mtRNAP. Ribbon representation of DNA, mtRNAP and TFB2M around the point of DNA melting in the LSP IC. Helices are depicted as cylinders. Functionally important residues in TFB2M are shown as sticks. The palm domain (residues 646-1230) of free mtRNAP (PDB ID: 3SPA; transparent pale green) (Ringel et al., 2011) was superimposed on the IC. For clarity, only residues 420-520 and 557-637 of mtRNAP are shown. Grey and green dashed lines indicate the trajectory of unresolved parts of the mtRNAP lever loop in the IC and apo mtRNAP structures, respectively. Arrows indicate the movement induced by TFB2M binding. (B) Nucleic acid binding by TFB2M. The surface of TFB2M is colored according to electrostatic potential with positive and negative potential in blue and red, respectively. DNA is shown in cartoon view. Upstream DNA and mtRNAP were omitted for clarity. (C) Activity of structure-based point mutants of mtRNAP and TFB2M in transcription assays.

Comparison with the free TFB2M structure reveals that the C-terminal tail of TFB2M (residues 389-396) has apparently moved to accommodate the intercalating hairpin of mtRNAP in the position observed in the IC (Figure 17). The tail may contribute to stabilizing the intercalating hairpin because deletion of eight amino acids from the C-terminus of TFB2M leads to a notable reduction in activity (Figure 16C).

Finally, TFB2M traps the non-template DNA strand in the open DNA region. This was previously suggested for Mtf1 (Paratkar and Patel, 2010) and is reminiscent, on the topological level, of the bacterial initiation factor sigma (Feklistov and Darst, 2011; Helmann and Chamberlin, 1988; Zhang et al., 2012). The NTD of TFB2M contributes an extensive positively charged surface that guides the DNA non-template strand away from the template strand (Figure 16B). Three conserved positively charged residues (R198, K201 and K202) protrude from the $\alpha 5$ helix of TFB2M towards the non-template strand and are required for efficient transcription initiation (Figure 16A and C, Figure 17). Further interactions may be mediated by the positively charged residues K153, R157, K163 and K206, which line the projected path of the non-template strand. Notably, three of these residues are located in the TFB2M-specific loop between $\beta 3$ and $\alpha 4$, possibly contributing to the specificity of TFB2M as a transcription initiation factor compared to TFB1M, which plays no role in transcription initiation (Litonin et al., 2010; Metodiev et al., 2009).

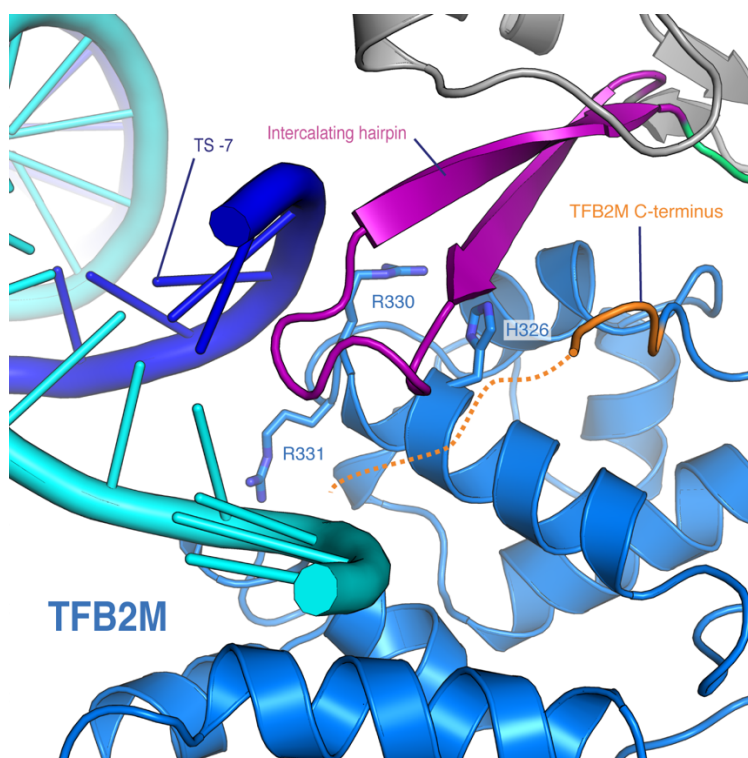


Figure 17. Structural basis of DNA melting and open DNA stabilization.

Close up view of the interaction of the TFB2M C-terminal domain with the intercalating hairpin of mtRNAP and the DNA. Nucleic acid and protein are depicted as ribbon representation with coloring as in Figure 12. The trajectory of the C-terminal tail of TFB2M (res. 393-396) observed in the TFB2M crystal structure but removed from the IC model is indicated as a dashed orange line. The presumably flexible tail would clash with the intercalating hairpin and the non-template DNA in the conformation observed in the free TFB2M structure.

Comparison to T7 RNAP initiation

Comparison of the mitochondrial IC structure to the structure of the T7 RNAP IC (Cheetham et al., 1999) reveals possible reasons for the requirement of initiation factors in mitochondria. Promoter recognition by T7 RNAP is achieved in part through sequence-specific DNA contacts at registers -5 to -11 formed by the specificity loop (Figure 18A and C). In contrast, the specificity loop in mtRNAP shows only fragmented density in the DNA major groove around registers -9 to -7, arguing against a prominent role in promoter recognition. Consistent with this, LSP and HSP share no sequence homology in this region and DNA base mutations hardly change initiation activity (Gaspari et al., 2004). In addition to the specificity loop, T7 RNAP engages with promoter DNA via the AT-rich recognition loop, which inserts into the upstream DNA minor groove between registers -17 and -13 (Cheetham et al., 1999). (Figure 18A and C). The structure of the IC demonstrates that mtRNAP does not form sequence specific contacts with promoter DNA in this region. Instead, only interactions between the PPR domain of mtRNAP and the upstream DNA backbone were detected (Figure 13A). Thus, recruitment of mtRNAP to DNA-bound TFAM apparently substitutes for the lack of extensive DNA interactions as compared to T7 RNAP. Opening of the DNA duplex by T7 RNAP is facilitated by the intercalating hairpin, which separates the two DNA strands at the upstream edge of the DNA bubble. In the apo mtRNAP structure, this element has been observed in a conformation incompatible with promoter melting (Ringel et al., 2011). In the mitochondrial IC, however, the intercalating hairpin and specificity loop are arranged as in the T7 RNAP IC (Figure 18B). This suggests that binding of TFB2M induces an initiation-competent conformation of mtRNAP (Figure 18A and B). In summary, comparison of the mitochondrial IC and the T7 RNAP IC suggests that TFAM compensates for the lack of prominent RNAP-promoter interactions upstream of the point of DNA opening and that TFB2M facilitates promoter opening by positioning a key structural element in mtRNAP in a fashion reminiscent of T7 RNAP.

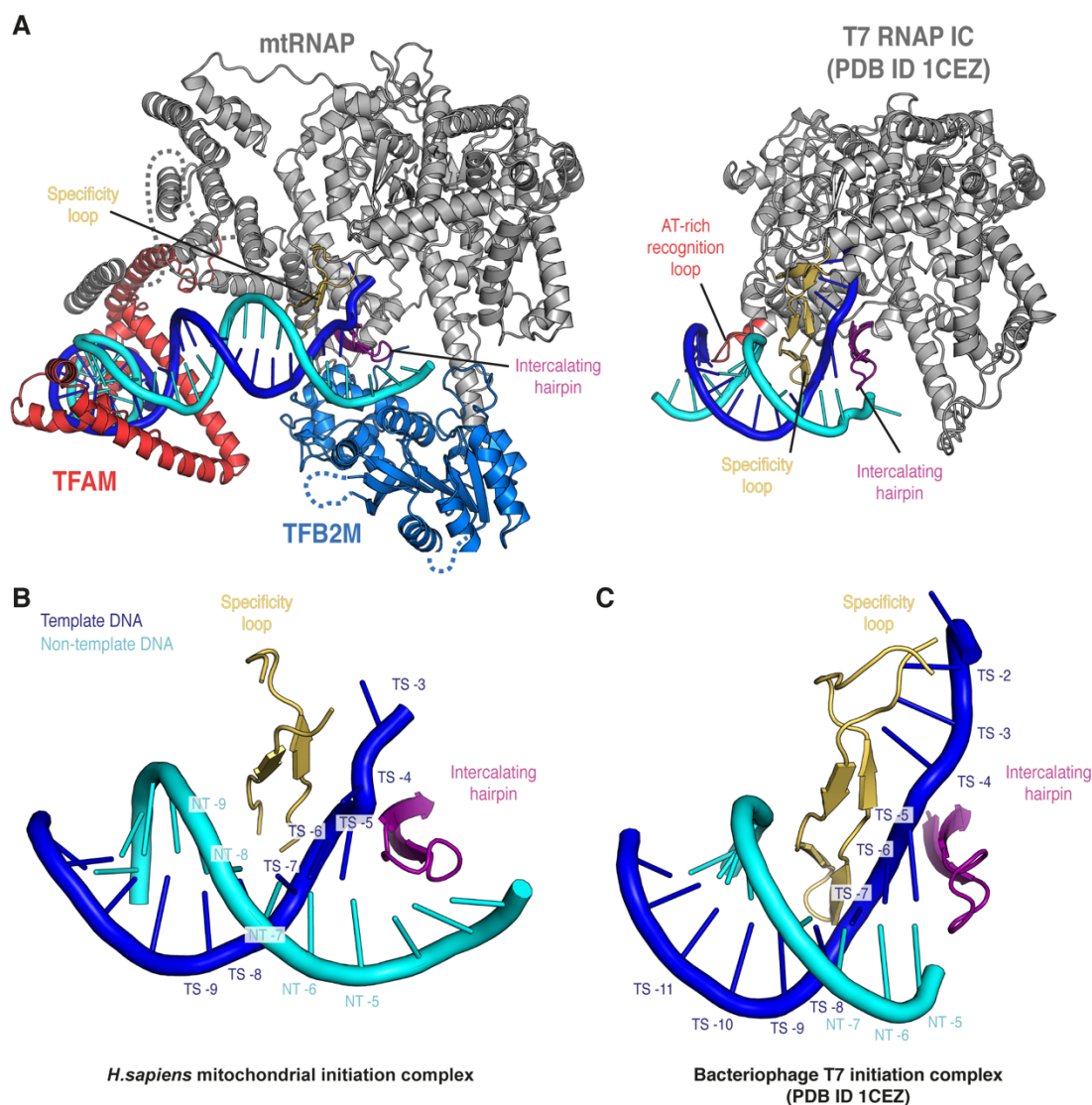


Figure 18. Comparison of human mitochondrial IC and T7 RNAP IC.

(A) (Left) Ribbon depiction of the LSP IC. MtRNAP is shown in grey with the intercalating hairpin in purple and the specificity loop in yelloworange. (Right) Ribbon depiction of the T7 RNAP initiation complex (PDB ID: 1CEZ) (Cheetham et al., 1999). Coloring as for the mitochondrial IC and with the AT-rich recognition loop in red. The topology around the point of DNA melting is similar in both complexes. (B) Ribbon representation of structural elements of human mtRNAP interacting with the DNA around the point of DNA melting. Coloring as in Figure 12 with base numbering indicated. The intercalating hairpin is positioned with the help of TFB2M to melt the DNA duplex and the specificity loop of mtRNAP runs along the DNA groove between bases -7 to -9. (C) Ribbon representation of the corresponding structural elements of T7 RNAP interacting with the DNA around the point of DNA melting in the T7 RNAP initiation complex (PDB ID 1CEZ) (Cheetham et al., 1999).

Transition from initiation to elongation

After RNA chain initiation, mtRNAP must lose its interactions with TFAM and TFB2M in order to transition to the elongation phase. In the case of T7 RNAP, this initiation-elongation transition is accompanied by substantial refolding of the polymerase, which destroys the PBD (Tahirov et al., 2002; Yin and Steitz, 2002). In contrast, comparison of the mitochondrial IC with the EC structure (Schwinghammer et al., 2013) demonstrates that the mtRNAP conformation remains largely unchanged and indicates rearrangements of the DNA during the initiation-elongation transition. In the EC, upstream DNA is repositioned and occupies the binding site of TFB2M, which must therefore dissociate during the transition (Figure 19). TFB2M dissociation is also required for binding of the elongation factor TEFM (Figure 19), as revealed in a recent structure of the EC bound by TEFM (Hillen et al., Cell, in press). The transition further creates a RNA exit channel underneath the intercalating hairpin, which remains in an open conformation and now separates the exiting RNA from the DNA template (Schwinghammer et al., 2013). Thus, TFB2M positions the intercalating hairpin for initiation, and this position is largely maintained during subsequent elongation. This concept of pre-organization of the polymerase conformation by an initiation factor has been observed for prokaryotic RNAP (Zhang et al., 2012). Taken together, these structural comparisons demonstrate the transitions that accompany promoter escape and highlight the distinct mechanisms employed by mtRNAP as compared to T7 RNAP.

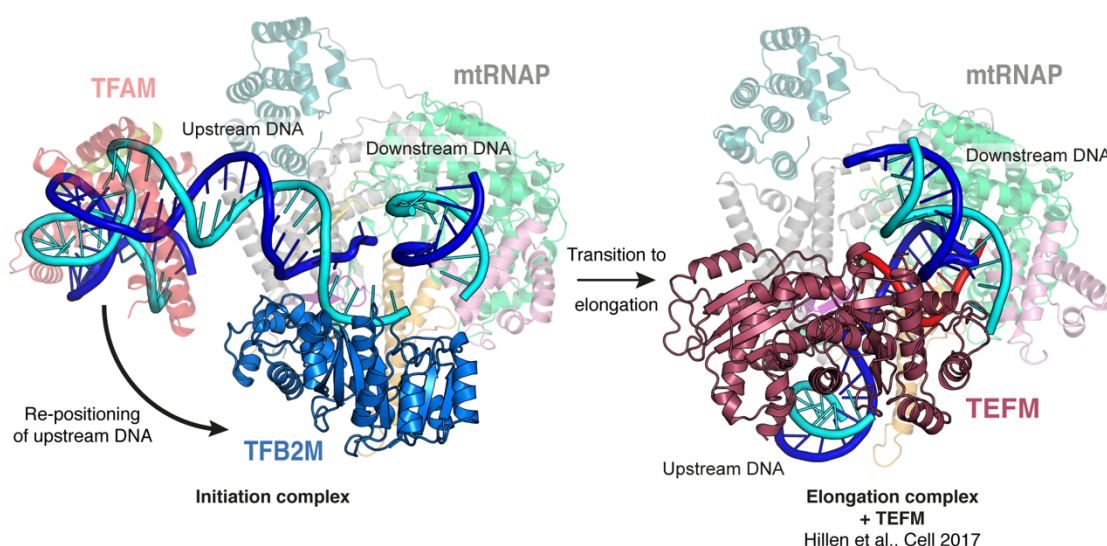


Figure 19. Transition from the initiation phase to the elongation phase of transcription.

(Left) Structure of the LSP IC. TFAM and mtRNAP are depicted transparently for clarity. The movement of the upstream DNA upon transition to the EC is indicated with an arrow. (Right) Structure of the human mitochondrial transcription elongation complex with the elongation factor TEFM bound (see chapter 3.2). MtRNAP is depicted transparently for clarity. The position of the downstream DNA duplex is identical in both the IC and the EC. The TFB2M binding site on mtRNAP is occupied by the upstream DNA and TEFM in the EC, demonstrating that a pronounced rearrangement of the upstream DNA must take place during the transition from initiation to elongation and that binding of TFB2M and TEFM to mtRNAP are mutually exclusive. The position of the intercalating remains in the open conformation.

3.1.3. Discussion

In this study we extend our previous structural work on the mitochondrial transcription from elongation (Schwinghammer et al., 2013) (see also section 3.2) to initiation mechanisms. Our structures of the ICs demonstrate the conserved architecture of the transcription complexes that assemble at divergent human mitochondrial promoters and support the sequential model of transcription initiation (Morozov et al., 2014) (Figure 20). First, recruitment of mtRNAP to TFAM-bound promoter DNA positions mtRNAP at the transcription start site. This explains the critical role of the distance between the TFAM-binding site and the start site in initiation (Dairaghi et al., 1995a). Subsequent binding of TFB2M induces DNA opening and stabilizes open DNA with the use of conformational changes and binding energy. Initial RNA synthesis may then be facilitated by the N-terminal region of TFB2M (residues 21-71), which is mobile in the IC structure but can be cross-linked to the initiating nucleotide (Sologub et al., 2009). Comparison of the IC with our recent EC structure bound to the mitochondrial elongation factor TEFM (Hillen et al., Cell, in press; see also section 3.2) demonstrates that TFB2M and TEFM binding to mtRNAP are mutually exclusive. The transition from initiation to elongation is accompanied by a dramatic re-arrangement of the upstream DNA instead of refolding of the polymerase, as observed for the distantly related T7 RNAP. These results not only provide the molecular basis for initiation, they also indicate changes that occur during the initiation-elongation transition and indicate that release of TFB2M is a prerequisite for TEFM recruitment to mtRNAP.

Our structural data show how the initiation mechanism of mtRNAP differs from that of phage T7 RNAP and multisubunit RNAPs. Whereas T7 RNAP does not depend on initiation factors, multisubunit RNAPs depend on additional factors for initiation, but these are not homologous to TFAM and TFB2M, neither on the sequence level nor the structural level. There are, however, conceptual and distant similarities between all initiation systems. In particular, mtRNAP uses its intercalating hairpin for DNA opening, like T7 RNAP, although the hairpin must be positioned by the initiation factor TFB2M. Also, TFB2M traps the non-template strand in the open DNA region, in a manner that is topologically similar to the sigma factor that is required by bacterial RNAP for initiation (Feklistov and Darst, 2011; Murakami and Darst, 2003; Zhang et al., 2012). Taken together, the transcription initiation system in mitochondria is unique, but some aspects of it show distant similarities with the bacteriophage and the bacterial systems.

In conclusion, our results illuminate the architecture of the IC and provide the structural basis of mitochondrial transcription initiation. The mitochondrial initiation system employs mechanisms of initiation that are clearly distinct from those observed for nuclear, bacterial or bacteriophage RNAPs. This likely reflects the need for regulating mitochondrial transcription, which is not only required for the expression of essential genes and synthesis of ribosomal and transfer RNA, but also to generate RNA primers for replication of the mitochondrial genome (Agaronyan et al., 2015; Gustafsson et al., 2016).

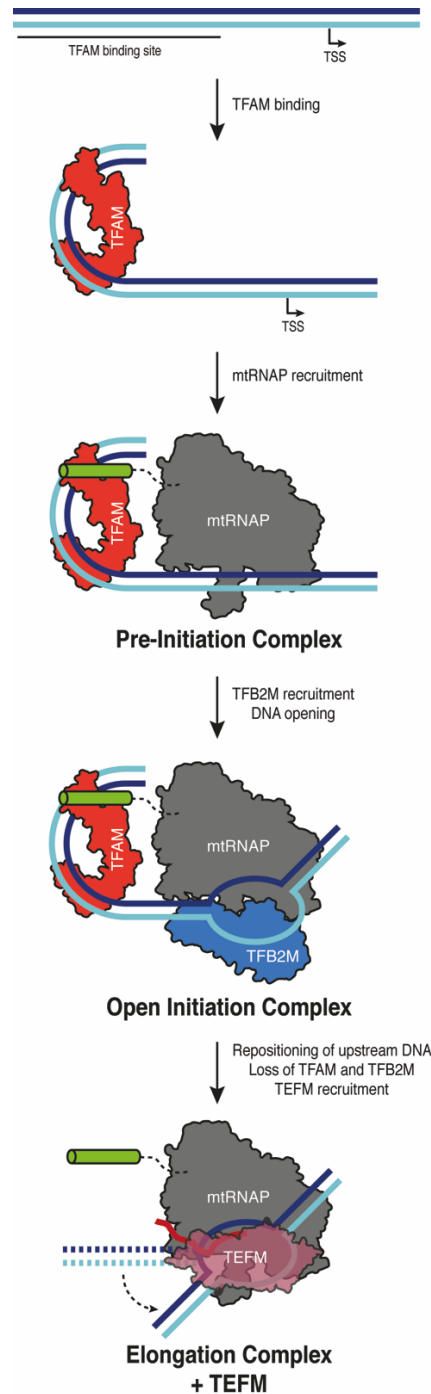


Figure 20. Model for transcription initiation in human mitochondria.

The TFAM binding site and the transcription start site (TSS) are indicated. TFAM (red) binding bends the DNA upstream of the TSS and leads to recruitment of mtRNAP (grey) to form the closed pre-initiation complex. Binding of TFB2M (blue) leads to melting of the DNA duplex and the bending of the downstream DNA observed in the open initiation complex. Promoter escape and transition to the elongation complex involve repositioning of the upstream DNA duplex (dashed arrow) and binding of TEFM (raspberry), which occupies a similar site on mtRNAP as TFB2M in the IC.

3.2. Mechanism of transcription anti-termination in human mitochondria

Results presented in this section were obtained in collaboration with the Temiakov lab at the Department of Cell Biology, Rowan University, Stratford, NJ, USA and with the Urlaub lab at the Max-Planck-Institute for Biophysical Chemistry and have been published:

H.S. Hillen, A.V. Parshin, K. Agaronyan, Y.I. Morozov, J.J. Graber, A. Chernev, K. Schwinghammer, H. Urlaub, M. Anikin, P. Cramer and D. Temiakov (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell*, in press

A detailed list of author contributions can be found on page VI.

For clarity, this results section presents all results of this publication. Structure determination and chemical cross-linking using BS3 and EDC was performed as part of this thesis work, while analysis of cross-links by mass spectrometry was performed by the Urlaub Lab and biochemical assays, pBpa and DSG cross-linking were performed by the Temiakov lab. Methods performed as part of this thesis work are described in section 2.4 and methods and additional data obtained predominantly by the collaborators are presented in section 5.2.1 and 5.2.2, respectively.

3.2.1. Abstract

In human mitochondria, transcription termination events at a G-quadruplex region near the replication origin are thought to drive replication of mtDNA by generation of an RNA primer. This process is suppressed by a key regulator of mtDNA – the transcription factor TEFM. We determined the structure of an anti-termination complex in which TEFM is bound to transcribing mtRNAP. The structure reveals interactions of the dimeric pseudonuclease core of TEFM with mobile structural elements in mtRNAP and the nucleic acid components of the EC. Binding of TEFM to the DNA forms a downstream "sliding clamp", providing high processivity to the elongation complex. TEFM also binds near the RNA exit channel to prevent formation of the RNA G-quadruplex structure required for termination and thus synthesis of the replication primer. Our data provide insights into target specificity of TEFM and mechanisms by which it regulates the switch between transcription and replication of mtDNA.

3.2.2. Results

TEFM contains two structured domains

Our initial crystallization trials did not produce crystals of full-length TEFM. Sequence based homology prediction suggested that human TEFM contains at least two regions with distinct structural organization (Minczuk et al., 2011), possibly interspersed by less well-ordered domains. Protease treatment is known to improve chances of crystallization of proteins with flexible regions (Dong et al., 2007). We therefore probed whether TEFM contains unstructured regions that may interfere with its crystallization by using limited proteolysis with trypsin, LysC and ArgC proteases. These experiments revealed two stable regions representing the N-terminal domain (NTD, ~10 kDa) and the C-terminal domain (CTD, ~26 kDa) of TEFM, while the region between these domains (the inter-domain linker) was sensitive to proteolysis (Figure 21A). Limited proteolysis of TEFM with ArgC prior to crystallization yielded large, well-diffracting crystals of the CTD (Figure 21B). The structure was subsequently solved at 1.9 Å resolution by single-wavelength anomalous diffraction (SAD) using selenomethionine-labeled protein (Table 16, Figure 21C). The TEFM CTD crystals belong to space group C2 and contain four molecules in the asymmetric unit with only minor differences between the individual molecules ($C\alpha$ RMSD < 1 Å over 200 residues). To obtain the crystal structure of the TEFM NTD, we expressed this domain separately and obtained large, rod-like crystals of space group $P4_32_12$ containing one molecule in the asymmetric unit (Figure 21B). The structure of the NTD was solved at a resolution of 1.3 Å using the anomalous signal of the naturally occurring sulphur atoms (native-SAD) (Table 16, Figure 21D). The crystal structures of the NTD and the CTD were refined to free R-factors of 19.7 and 22.8, respectively, and both showed excellent stereo chemical quality (Table 16).

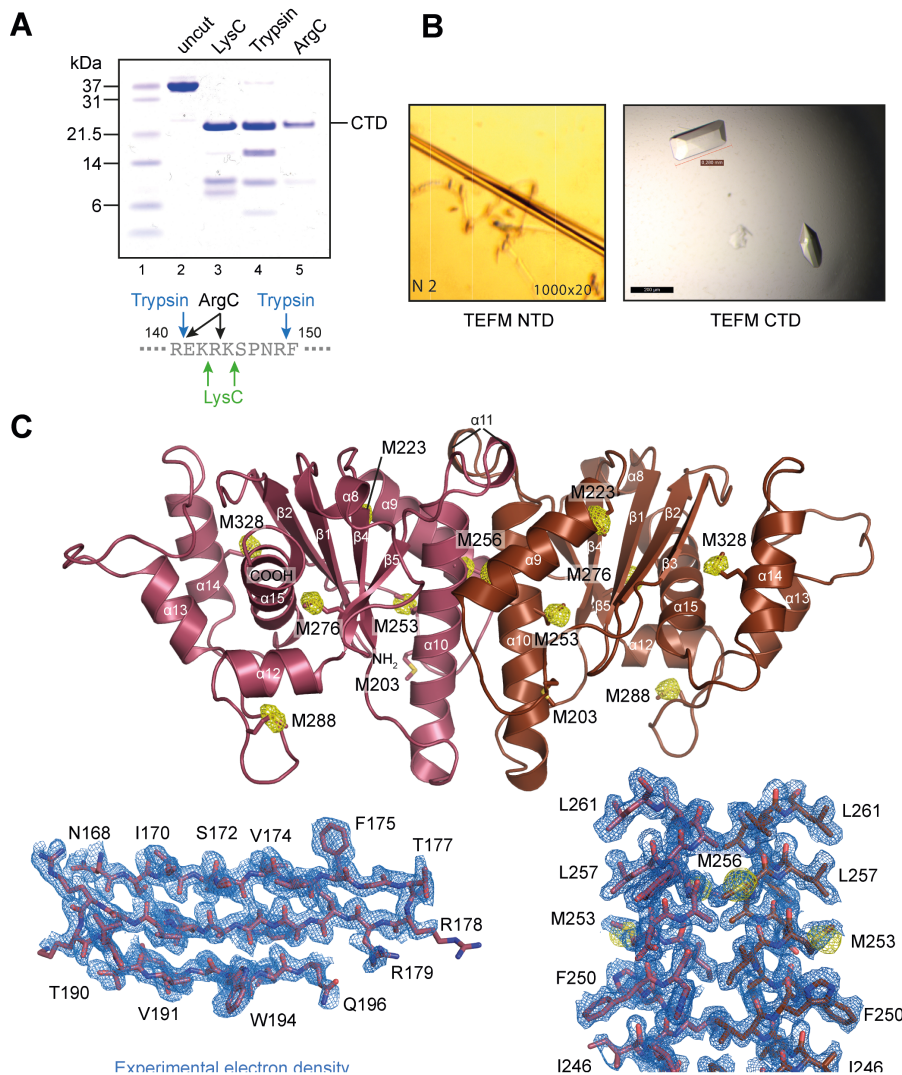


Figure 21. Domain organization, crystallization and structure determination of human TEFM.

(A) Limited proteolysis of TEFM. (Top) Cleavage of TEFM with proteases occurs in the inter-domain linker. (Bottom) Schematics of cleavage in the inter-domain linker of TEFM as determined by mass spectrometry. (B) (Left) Exemplary optimized crystal of the TEFM N-terminal domain. (Right) Exemplary optimized crystals of the TEFM C-terminal domain. (C) Structure determination of the TEFM CTD. (Top) Ribbon representation of the structure of the TEFM CTD with one monomer colored in raspberry and the other in chocolate. Secondary structure elements are indicated. Methionine residues substituted with selenomethionine in the crystals used for phasing are indicated and shown as sticks. An anomalous difference Fourier map calculated using the dataset used for phasing is shown as mesh at 5 σ in yellow. (Bottom) The unbiased experimental electron density used for building the initial model is

shown as mesh (marine blue) at 1 σ around representative regions shown as sticks (sheets β 1, β 2, β 3 and the dimerization interface composed of α 10 from each monomer). **(D)** Structure determination of the TEFM NTD. Left. Ribbon representation of the structure of the TEFM NTD colored in orange. Secondary structure elements are indicated. Sulphur-containing residues M116 and C130 are indicated and shown as sticks and an anomalous difference Fourier map calculated using the dataset used for phasing is shown as mesh at 5 σ in pink. Right. Unbiased experimental electron density used for building initial model is shown as marine blue mesh at 1 σ around representative helix α 5.

Structure of TEFM

In the two crystal forms, we observed interpretable electron density for residues 57-134 (NTD) and residues 153-356 (CTD), which comprise 87% of the full-length, mature TEFM (Table 16). Our structures lack 21 N-terminal residues, 3 C-terminal residues, and a 19-amino acid linker region, which connects the two domains. The linker contains six lysine and two arginine residues and appears to be disordered in the CTD structure. The TEFM CTD forms a dimeric assembly of monomers containing an RNase H-like fold that is characteristic of members of the Holliday junction resolvase family of enzymes (Figure 22B and Figure 23) (Wyatt and West, 2014). This family includes a large number of RuvC-like resolvases from bacteria, viruses, and mitochondria (Ariyoshi et al., 1994; Górecka et al., 2013; Roe et al., 1998). Structural alignment suggests that the closest structural homolog of TEFM CTD is *Schizosaccharomyces pombe* mitochondrial Holliday junction resolvase, Cce1 (Ydc2) (Ceschini, 2001). While TEFM and Cce1 share little in sequence homology, the structures contain a common core fold consisting of five β -strands and four C-terminal helices (Figure 22D). Compared to Cce1, TEFM lacks a protruding three-helix region, and two N-terminal helices and two loops (α 12- α 13; α 13- α 14) are expanded (Figure 22D) (Sigala and Tsaneva, 2003). Similar to Cce1, dimerization of the TEFM CTD is mediated by a hydrophobic interface mainly formed by a contact between the α 10 helices (residues 244–266) from the two CTD monomers.

A unique feature of TEFM, not observed thus far in any related bacterial or mitochondrial Holliday junction resolvase, is the presence of a compact globular N-terminal domain (Figure 22B and Figure 23). The structure of the NTD reveals that this domain adopts a tandem helix-hairpin-helix fold ($[HhH]_2$) found in a number of DNA binding proteins. HhH domains have been implicated in non-sequence-specific double stranded DNA binding and have been found in bacterial and eukaryotic transcription factors and other DNA binding proteins (Close et al., 2011; Johnson et al., 2008). Although lacking apparent sequence homology, the TEFM NTD superimposes well with the HhH domain of the bacterial Tex protein (C α RMSD = 1.4 Å over 56 residues, PDB ID: 3BZC, (Johnson et al., 2008)) (Figure 22E). Taken together, TEFM exhibits a previously unseen combination of distinct structural domains – the Holliday junction resolvase-like CTD core flanked by two globular HhH domains.

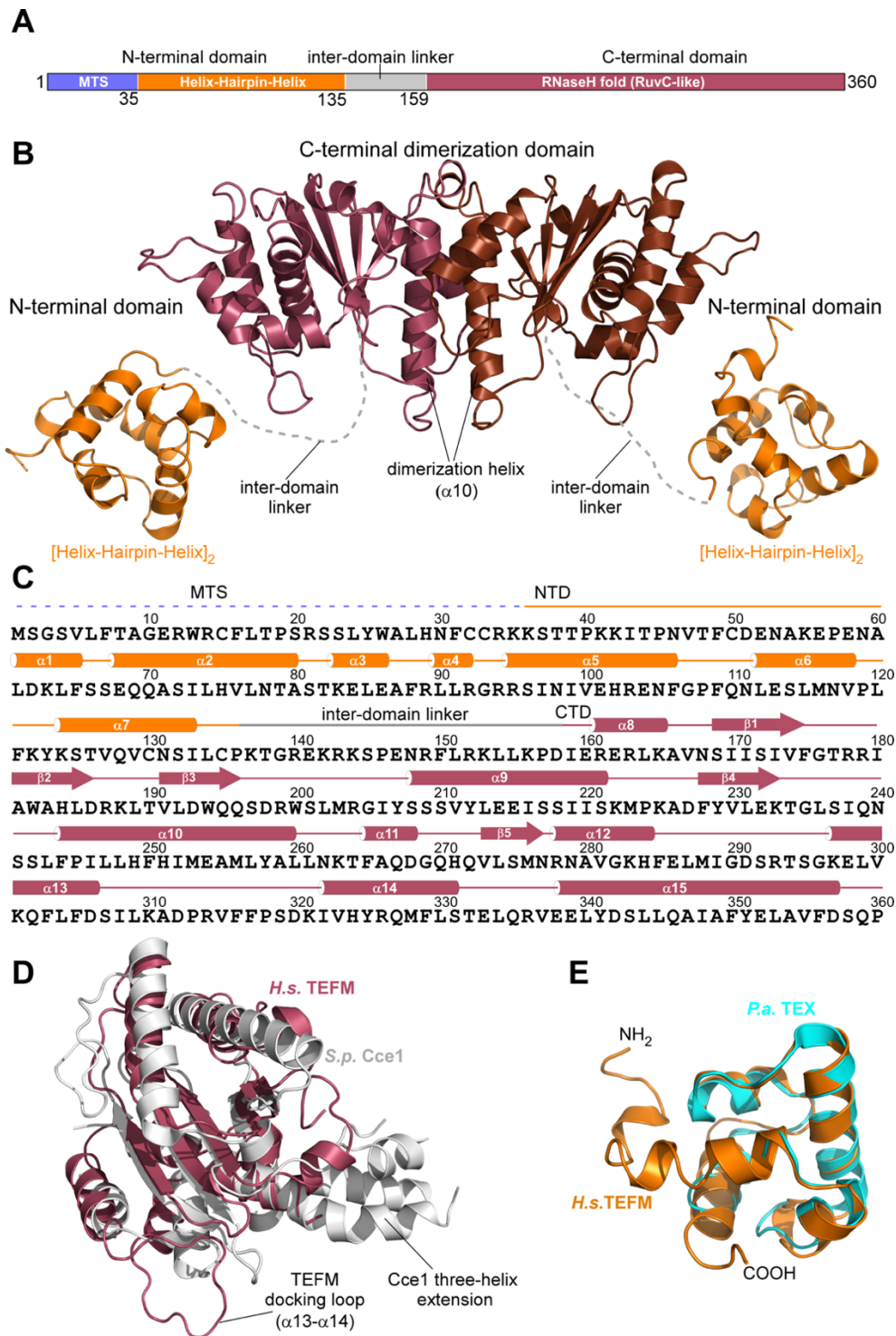


Figure 22. Crystal structure of human TEFM.

(A) Schematic representation of human TEFM with residue numbers indicated. Domains identified by sequence-based homology prediction are depicted in different colors: mitochondrial targeting sequence (MTS): blue; N-terminal domain (NTD): orange; C-terminal domain (CTD): raspberry, inter-domain linker: grey. (B) A ribbon model of TEFM with the major domains and structural elements indicated. The C-terminal dimeric domain (CTD) is in brown/raspberry, the N-terminal domains (NTD) in orange. The inter-domain linker is represented by a dashed, grey line. (C) Schematics of the primary and secondary structures of TEFM. (D) Structural alignment of TEFM CTD (raspberry) and fission yeast Holliday junction resolvase Cce1 (PDB ID: 1KCF) (Ceschini, 2001) (grey). (E) Structural alignment of TEFM NTD (orange) and bacterial (*P. aeruginosa*) transcription factor TEX (PDB ID: 3BZC) (Johnson et al., 2008)(cyan).

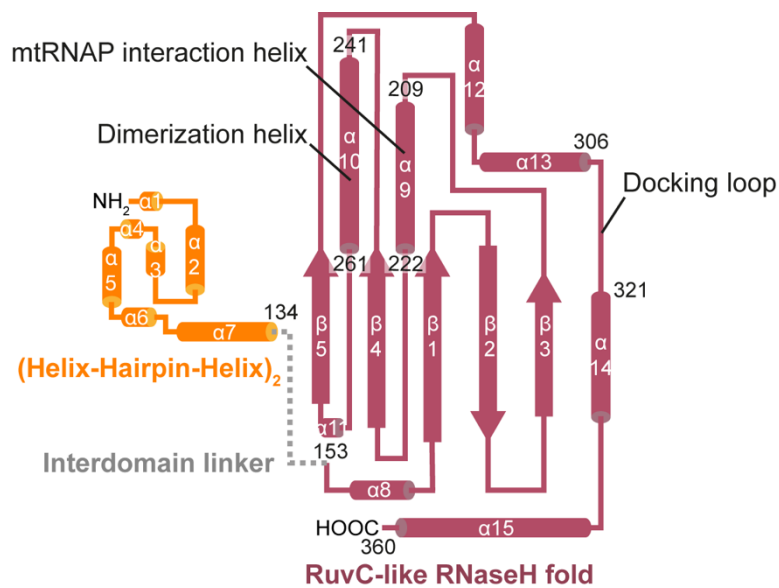


Figure 23. Secondary structure topology of human TEFM.

Schematic representation of secondary structure elements in TEFM. Alpha-helices are shown as cylinders, β -sheets as arrows. Important structural features are indicated. Coloring as in Figure 22A.

Functional role of TEFM domains

TEFM has been previously implicated in anti-termination activity and was shown to affect mtRNAP stability and processivity (Agaronyan et al., 2015; Minczuk et al., 2011; Posse et al., 2015). To determine the functional role of TEFM domains in these activities, we constructed variants representing the CTD alone ($\Delta 159$ TEFM, residues 160-360), the CTD with a part of the linker ($\Delta 144$ TEFM, residues 145-360) and the CTD with the linker ($\Delta 135$ TEFM, residues 136-360) (Figure 24A). In the anti-termination assay, transcription of template DNA containing the CSBII region results in termination of RNA synthesis by mtRNAP in the absence of TEFM, whereas its addition results in accumulation of run-off products (Figure 24B, lanes 1-4). We found that the NTD of TEFM is not required for anti-termination activity (Figure 24B, lanes 5-7). Removal of 10 amino acids of the linker along with the NTD results in significant loss of anti-termination activity (Figure 24B, lanes 8-10), while deletion of the entire linker subdomain abolished anti-termination (Figure 24B, lanes 11-13). TEFM variants containing single or multiple substitutions in the linker were defective to various extents in anti-termination activity, and substitution of amino acid residues 149-153 resulted in loss of activity (Figure 24C and D). We therefore conclude that the inter-domain linker is required for anti-termination activity of TEFM.

We next analyzed which part of TEFM affects its ability to stabilize the EC (Figure 24E). In this assay, ECs are halted 35 bp downstream from a promoter (in the absence of CTP) for 60 min prior to a “chase” to run-off transcription. In the absence of TEFM, halted ECs are not stable and accumulation of 35-mer RNA product is observed (Figure 24E, lane 1). Addition of TEFM, however, results in stabilization of the EC, prevents high mtRNAP turn-over and, consequently, causes nearly complete extension of 35 nt RNA to run-off (Figure 24E, lanes 2-

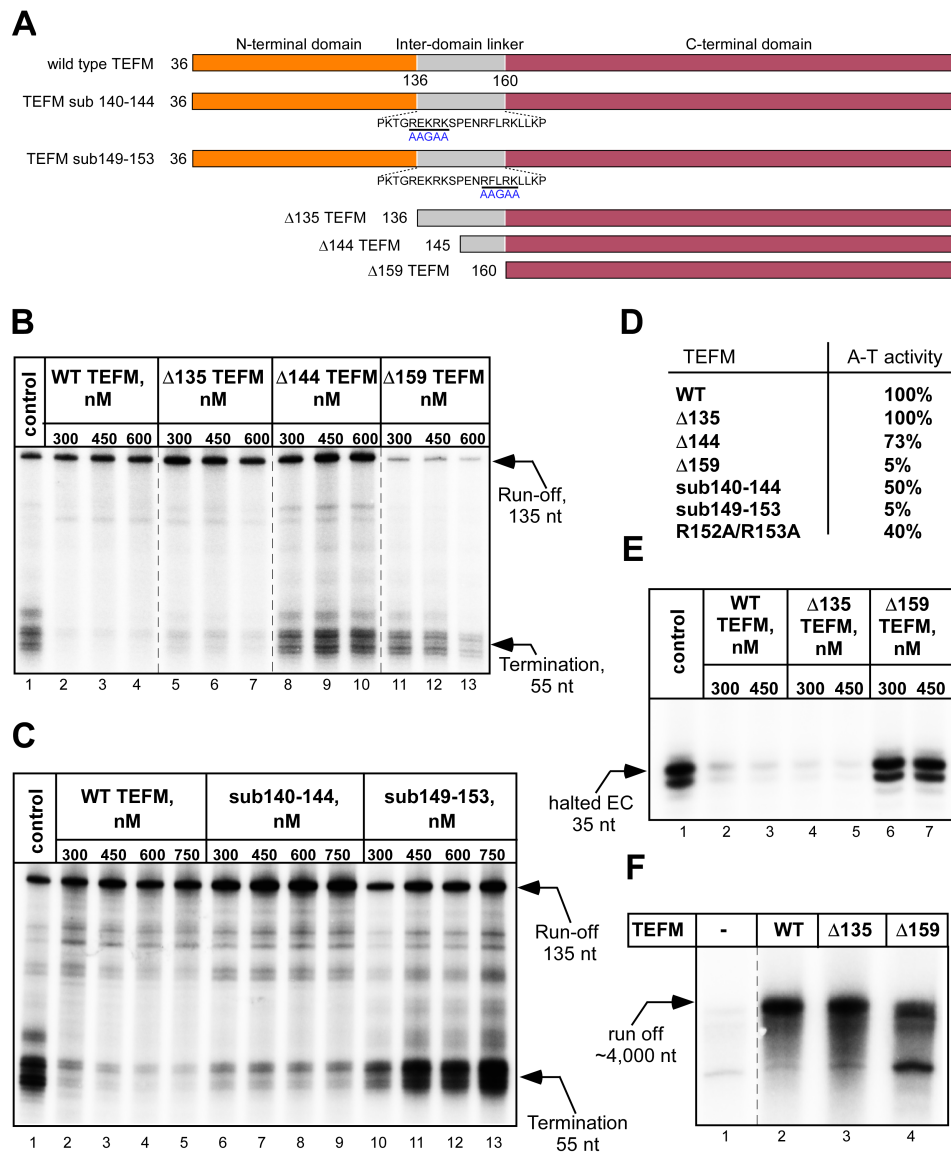


Figure 24. Functional role of TEFM domains.

(A) Schematic representation of the TEFM NTD truncation mutants and inter-domain linker substitution mutants used. Residues substituted are underlined with a black bar and the replacing residues are shown in blue beneath. (B) TEFM CTD with the linker are required for anti-termination activity. Transcription assays were performed using a PCR template containing LSP and the CSBII region. The control reaction (lane 1) contains no TEFM. The reaction products were resolved using 20% PAGE containing 6M Urea. (C, D) Anti-termination activity of TEFM variants having substitutions in the inter-domain linker. (E) The NTD of TEFM is not required for the EC stability. The ECs assembled in the presence or absence of TEFM variants were halted 35 nt downstream of the transcription start site by omitting CTP. Upon incubation, the complexes were chased with CTP and the products of the reaction resolved using 20% PAGE containing 6M Urea. The region of the gel containing the 35 nt RNA is shown. (F) Only the CTD of TEFM is required for mtRNAP processivity. Transcription was performed using a linearized plasmid template containing the LSP promoter to generate ~4000 nt run-off product.

3). As in the experiments described above, deletion of the NTD did not affect TEFM stabilization activity (Figure 24E, lanes 4-5), while loss of both NTD and inter-domain linker rendered TEFM incapable of EC stabilization (Figure 24E, lanes 6-7). Finally, we examined whether the TEFM variants were able to support mtRNAP processivity on long templates

(Figure 24F). In the absence of TEFM, mtRNAP is not able to produce detectable 4,000 nt long run-off RNA products. In contrast, run-off synthesis was observed in the presence of TEFM or TEFM variants lacking the NTD (Figure 24F, lanes 2-3). In the presence of TEFM CTD, mtRNAP was still able to generate long RNA transcripts, suggesting that the CTD, even without the inter-domain linker, is sufficient for stimulating processivity (Figure 24F, lane 4).

Structure determination of TEFM-mtRNAP elongation complex

To obtain the crystal structure of an anti-termination complex, in which elongating mtRNAP is bound to TEFM, we first defined the topology of the nucleic acid scaffold in the EC by RNase I and DNase I foot-printing (Figure 25A). In the absence of TEFM no clear RNA footprint was detected, likely due to the relative instability of the halted ECs. In contrast, in the presence of TEFM, a distinct footprint of about 19 nt of protected RNA was observed suggesting that TEFM stabilizes the EC by interacting with the 5' end of the exiting transcript (Figure 25A). Removal of the NTD did not significantly alter the TEFM footprint, identifying the CTD and the linker domain as primary RNA-interacting regions (Figure 35A and B).

TEFM has previously been shown to interact with the DNA in the EC (Agaronyan et al., 2015). To determine the TEFM footprint on DNA, we assembled ECs using an RNA-DNA scaffold and probed TEFM protection by DNase I (Figure 25B). In the presence of TEFM, the protected region was significantly larger than for the EC alone, and included about 12 bp of upstream DNA and 14 bp of downstream DNA (Figure 25B and C). We did not detect any changes in DNA protection when the $\Delta 135$ TEFM variant was used, suggesting that the NTD does not strongly associate with DNA (Figure 35C). However, a TEFM variant lacking both NTD and the linker did not produce a clear DNA foot-print (Figure 35D). Overall, binding of TEFM to the EC significantly increases the region of nucleic acid protection (Figure 25C), consistent with our previous findings (Agaronyan et al., 2015).

To crystallize the anti-termination complex, we used the $\Delta 135$ variant of TEFM, which is fully functional in anti-termination assays, and $\Delta 150$ mtRNAP assembled into an EC on an RNA-DNA scaffold derived from the foot-printing experiments above. Crystals of the EC-TEFM complex belonged to space group C2, contained two copies of the EC-TEFM complex in the asymmetric unit and diffracted to 3.9 Å resolution. The structure was solved by molecular replacement and could be refined to a free R-factor of 27.6% (Table 17, Methods).

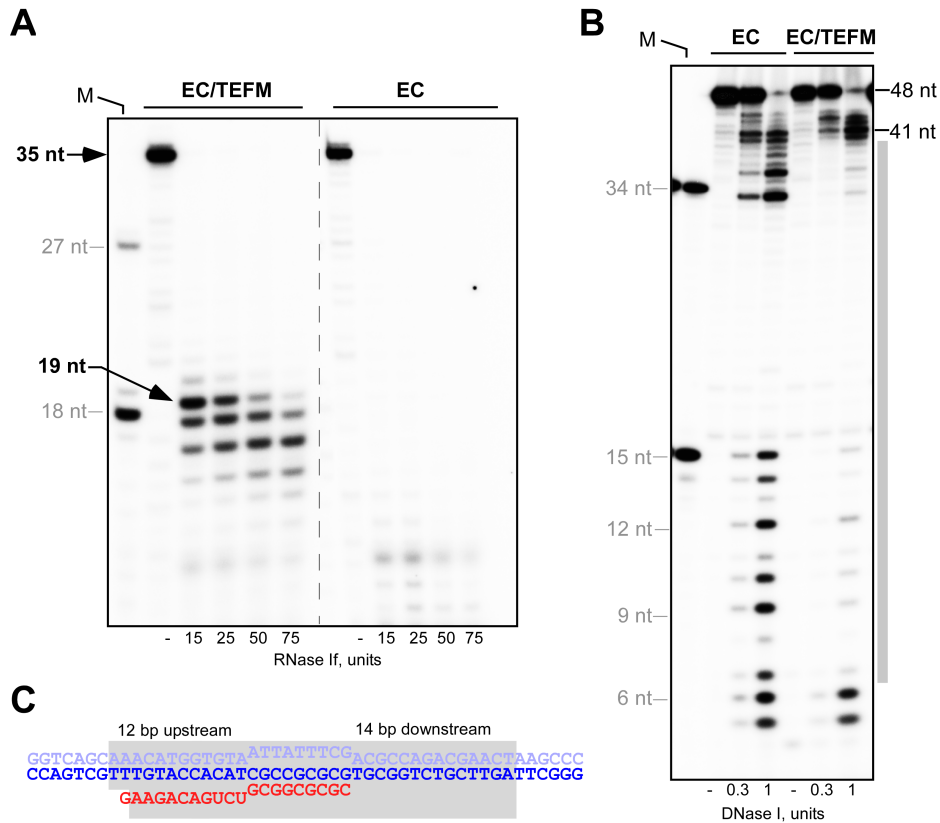


Figure 25. Topology of the anti-termination complex.

(A) TEFM protects 19-20 nt of RNA from RNase I cleavage in the EC-TEFM complex. The ECs were halted 35 bp downstream of the LSP promoter prior to RNase I treatment. Positions of the RNA markers (18 and 27 nt) are indicated by grey lines. (B) TEFM protects the downstream and the upstream DNA in the EC. The EC was assembled using R14/TS26/NT26A scaffold (see 5.2.1) in the absence or presence of TEFM prior to DNase I treatment. The length of the DNA markers (15 and 34 nt, the leftmost lane) is indicated. The grey bar represents the DNA region protected by TEFM. (C) Schematics of the RNA and DNA protection regions (shown in grey) in the anti-termination complex shown on scaffold used in (B).

Architecture of the TEFM-mtRNAP elongation complex

The structure revealed that the TEFM CTD dimer forms an extensive network of interactions with all components of the EC by binding mtRNAP at the open side of the RNA-DNA hybrid cavity (Figure 26A and B). The primary point of interaction between mtRNAP and TEFM is the intercalating hairpin (Figure 26C), a mobile element in the N-terminal domain of mtRNAP involved in separation of RNA from the RNA-DNA hybrid and thus maintenance of the trailing edge of the transcription bubble (Schwinghammer et al., 2013). Two invariant aromatic residues in the intercalating hairpin, Y610 and F612, and the conserved residue Q617 project into a hydrophobic cleft on the surface of TEFM, the walls of which are formed by helices adjacent to the β -strand core (Figure 26C).

The structure of the EC-TEFM complex reveals that the homodimeric form of TEFM observed in the CTD crystals also occurs when TEFM is bound to mtRNAP (Figure 26A and B). To investigate whether dimerization of TEFM is required for its activity, we modified the hydrophobic interface between the TEFM monomers to prevent their association by

substituting five residues in the $\alpha 10$ helix (Figure 27A). In contrast to the wild type TEFM, the resulting protein eluted as a monomer during size exclusion chromatography and did not bind to the EC (Figure 27B). No anti-termination activity was observed with the mutant monomeric TEFM, confirming that dimerization of TEFM is required for its activity (Figure 27C).

TEFM makes extensive interactions ($\sim 1000 \text{ \AA}^2$) with the downstream DNA duplex of the EC. The positively charged surface of one of the TEFM monomers covers the otherwise exposed DNA and along with mtRNAP sequesters it in a fashion resembling the DNA "clamping" observed in multi-subunit eukaryotic and bacterial RNAPs (Figure 26D, Figure 28A and B). This clamping may explain the stabilization and higher processivity of ECs in the presence of TEFM (Agaronyan et al., 2015).

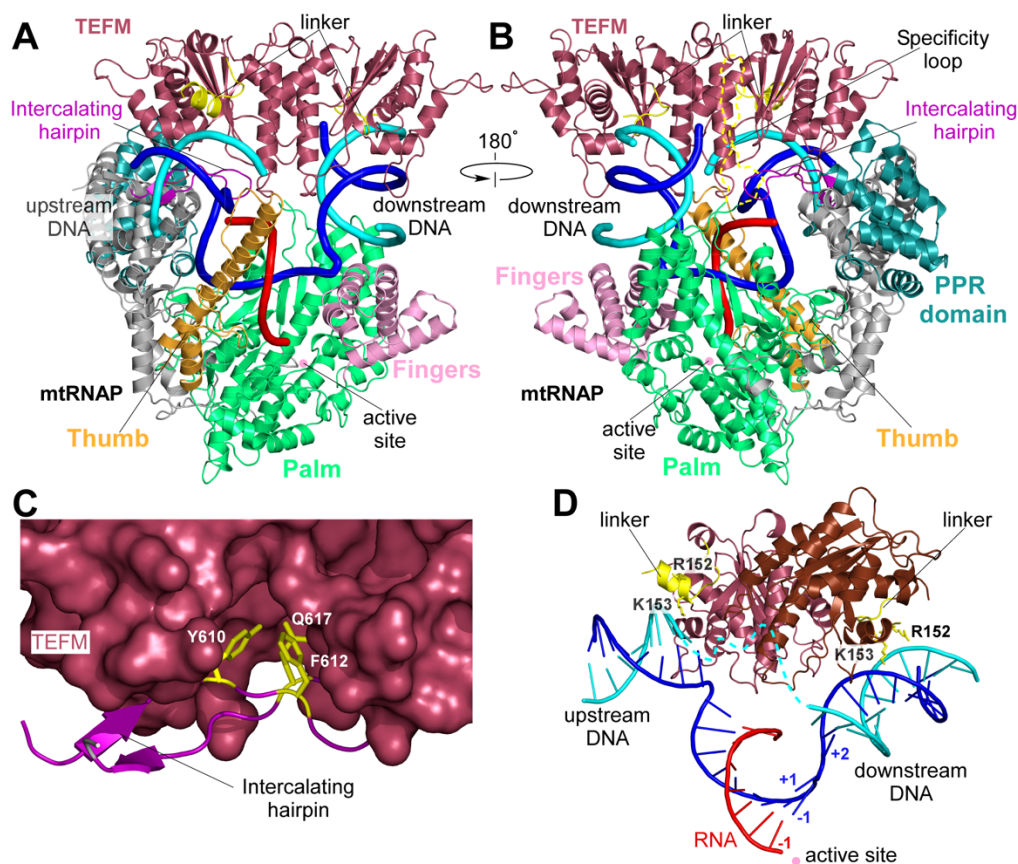


Figure 26. Structure of the anti-termination complex.

(A,B) TEFM binding to the EC. MtRNAP (ribbon model) is shown with the major domains colored: Thumb: bright orange, Palm: lime-green, NTD: grey, PPR: teal, Fingers domain: pink. RNA is shown in red, template DNA strand in blue, non-template DNA strand in cyan. Intercalating hairpin and specificity loop (modeled) of mtRNAP are highlighted in purple and yelloworange, respectively. (C) Close-up view of the TEFM binding site at the intercalating hairpin of mtRNAP. TEFM is shown as surface representation. Conserved residues (yellow) in mtRNAP are shown as sticks. (D) TEFM interacts with the RNA-DNA scaffold in the EC. TEFM (ribbon representation) binds at junctions of the RNA-DNA hybrid and interacts with downstream and upstream DNA. The +1 template DNA base is not paired to its non-template counterpart, indicating the post-translocated conformation state of the EC. Note the proximity of the inter-domain linkers (residues 149-159, yellow) to the DNA duplexes; positively charged residues implicated in interaction with DNA are shown as sticks.

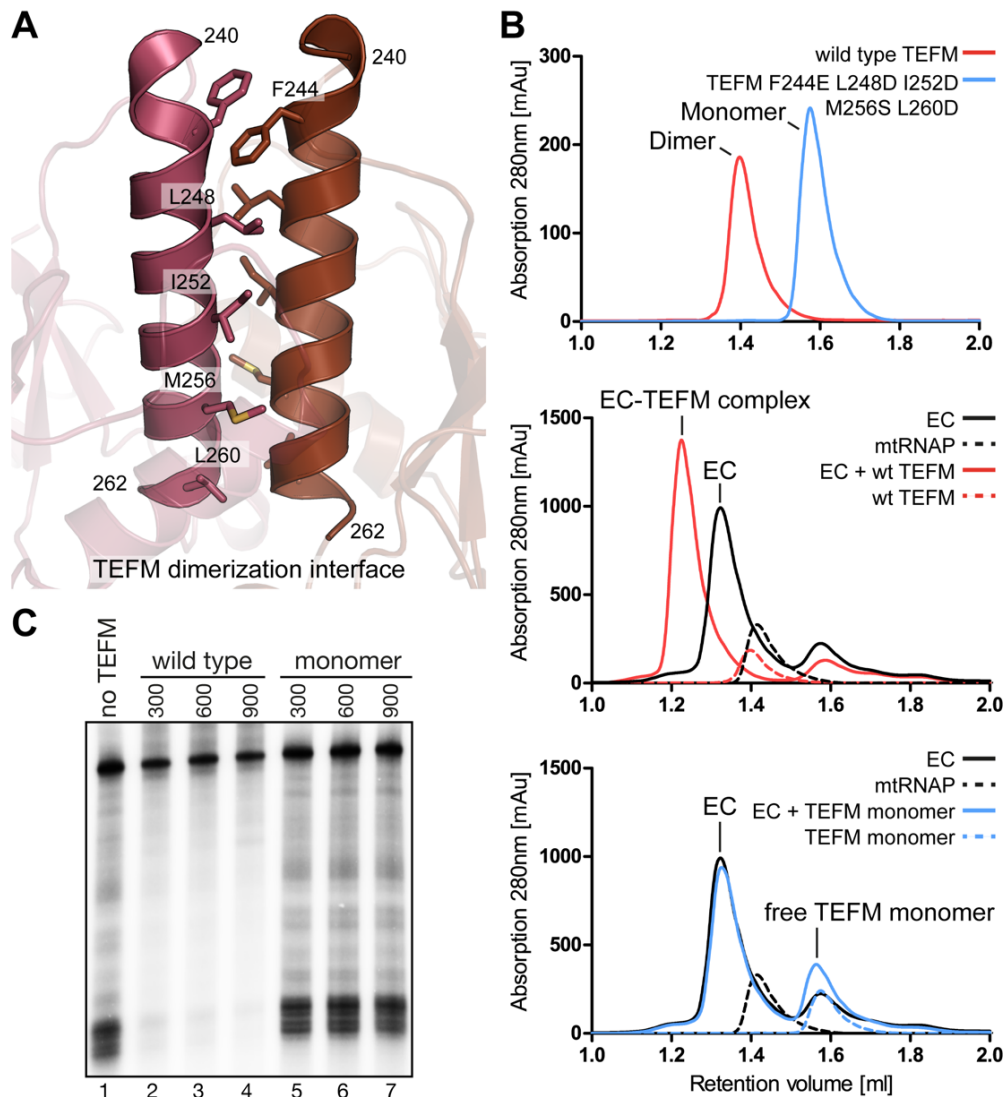


Figure 27. Dimerization of TEFM is required for EC binding and anti-termination activity.

(A) Close-up view of the dimerization interface of TEFM. Coloring and representation as in Figure 1B. The $\alpha 10$ helix of each monomer is highlighted, with the rest of the protein shown transparently. Hydrophobic residues forming the dimerization interface that were subjected to mutagenesis are shown as sticks. (B) Mutation of the TEFM residues contributing to the dimerization interface results in a monomeric form of TEFM that cannot bind EC. Upper panel: A monomeric TEFM mutant ($\Delta 35$ TEFM F244E L248D I252D M256S L260D) was purified and used for analytical size exclusion experiments. The mutant TEFM elutes as a monomer (blue), while the WT TEFM elutes as a dimer (red). Middle panel: Incubation of the EC with WT TEFM leads to a distinct shift of the peak representing the EC, indicating formation of a stoichiometric complex (2:1 TEFM:EC). In contrast, incubation of the EC with the monomeric TEFM mutant does not lead to a change in retention volume as compared to the individual components (Lower panel). (C) The monomeric TEFM mutant does not prevent termination at CSBII. Transcription assays were performed using the PCR template containing LSP and the CSBII region. The control reaction (lane 1) contains no TEFM. The reaction products were resolved using 20% PAGE containing 6M Urea.

The positively charged TEFM surface extends to the other monomer contributing to a much smaller interaction surface with the upstream than with the downstream DNA (Figure 28A and B). Consequently, the electron density for the downstream DNA duplex was strong, whereas the density for the upstream DNA duplex was weaker due to higher conformational flexibility

(Figure 28C). Furthermore, we observed difference density corresponding to the single-stranded region of the DNA non-template strand in the transcription bubble, which wraps around the opposite side of the dimer axis of TEFM formed by the $\alpha 10$ helices (Figure 28C). This region of the DNA was invisible in the previous structure of the EC (Schwinghammer et al., 2013) due to conformational flexibility but seems to be stabilized in the EC-TEFM complex. However, we refrained from modelling it due to the limited resolution of the complex structure. Thus, in addition to interacting with the downstream DNA, TEFM also stabilizes the EC by sequestering the single-stranded portion of the non-template strand of DNA (Figure 26D). Although visible only partially in the EC-TEFM structure, both inter-domain linkers of TEFM are located in close proximity to the DNA suggesting that the NTDs of TEFM may reside near or at the downstream and upstream duplexes (Figure 26D).

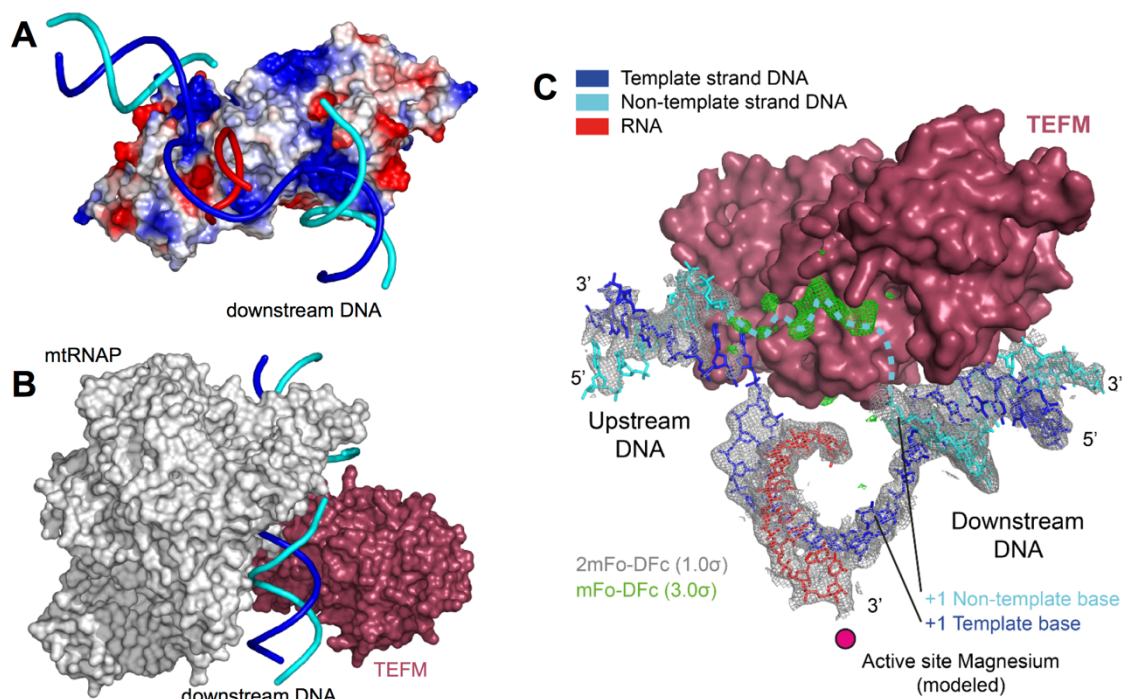


Figure 28. Properties of the anti-termination complex.

(A) Surface charge distribution in TEFM. Surface charge analysis of TEFM reveals an extensive positively charged surface facing the nucleic acid scaffold in the EC-TEFM complex. (MtRNAP is not shown). (Blue: positive charge; red: negative charge). DNA is shown in cartoon view with coloring as in Figure 26A. Vacuum electrostatic potential was calculated with Pymol. (B) TEFM interactions with the downstream DNA contribute to the formation of the "sliding clamp". TEFM (raspberry) and mtRNAP (grey) are shown in surface representation. (C) Electron density for nucleic acids in the anti-termination complex. The final model of the non-template DNA, template DNA and RNA is shown as sticks in cyan, blue and red, respectively. The TEFM CTD is shown as surface representation in raspberry. The 2mFo-DFc electron density for the nucleic acid included in the model is shown as grey mesh at 1 σ . An unbiased mFo-DFc map for the region corresponding to the single-stranded portion of the non-template strand (not modeled) is shown as green mesh at 3 σ . The expected trajectory of the missing non-template bases is indicated by a dashed line.

Thus, it is plausible that the positively charged residues in the linker contribute additional interaction points with the upstream and downstream DNA, thereby anchoring TEFM to the EC (Figure 26D). This could be important for the anti-termination activity because binding of TEFM to the EC must outweigh the energetically favorable formation of stable secondary structure such as a G-quadruplex. Indeed, substitutions of the charged residues in the linker region or its deletion have a profound effect on transcription anti-termination as demonstrated above (Figure 24).

Implication for mtRNAP translocation

Binding of TEFM to the EC does not cause noticeable structural changes in the elongation factor ($C\alpha$ RMSD = 0.48 Å over 389 residues). Superimposition of mtRNAP in the EC and the EC-TEFM complex reveals that the C-terminal catalytic domain remains unchanged, whereas the PPR domain and the N-terminal core domain are slightly rotated outwards and the "fingers" domain adopts a markedly different orientation (Figure 26A, Figure 29A). The latter, however, is not caused by TEFM binding *per se*, as the observed changes are attributed to the different translocational state of mtRNAP. Indeed, the "n+1" template strand base in the EC-TEFM complex is not paired to the corresponding non-template strand base and instead is found in a "pre-insertion" site of mtRNAP, indicative of the post-translocated state of the EC (Figure 26D) (Temiakov et al., 2004). The O/Y helices of the fingers domain are rotated $\sim 15^\circ$, respectively, as compared to the pre-translocated complex (Schwinghammer et al., 2013), and thus the active site of mtRNAP is found in the "open" conformation, previously observed in T7 RNAP ECs (Figure 29B and C) (Tahirov et al., 2002; Yin and Steitz, 2002). Overall, the conformation of the fingers domain and the position of the 3' end of RNA in mtRNAP closely matches those in the T7 RNAP post-translocated complex. Comparison of the available structures of mtRNAP (the apo form, pre-translocated EC and post-translocated EC) suggests a rotation of the O/Y helices around an axis that is nearly parallel to the O helix (Figure 29D). This rotation allows for delivery of the cognate substrate bound to the n+1 template base into the insertion site and the closing of the active site required for catalysis (Temiakov et al., 2004; Yin and Steitz, 2004). The mechanism of translocation therefore appears to be highly conserved between mitochondrial and phage RNAPs.

Considering the conformational changes in mtRNAP observed in the EC-TEFM structure, we employed a pyrophosphorolytic assay to probe whether TEFM can affect the translocation state of the EC (Figure 29E). Pyrophosphate (PPi) binds only to ECs in which mtRNAP is in a pre-translocated conformation and drives a reaction that is reverse to RNA synthesis. In the absence of TEFM, RNA was readily shortened by 2 nucleotides, suggesting pyrophosphorolytic activity of this EC (Figure 29E, lanes 3-7). In contrast, in the presence of TEFM the sensitivity of the EC was significantly decreased suggesting stabilization of the post-translocated state, consistent with our structural observations (Figure 29E, lanes 9-13).

translocated EC (PDB ID: 4BOC) (Schwinghammer et al., 2013) and post-translocated EC-TEFM (this work) were superimposed using the conserved palm domain (residues 643-1230). (E) TEFM stabilizes the post-translocated state of the EC. A pyrophosphorolytic assay was performed using EC (lanes 3-7) and EC-TEFM (lanes 9-13). Control (lane 1) - the EC. The assembled complexes were nearly 100% active as evident by efficient extension of the RNA primer by GTP (lanes 2,8).

Mapping of TEFM-mtRNAP interactions by cross-linking

Since interactions between mtRNAP and TEFM may involve structural elements that are mobile and thus only partially visible in the EC-TEFM structure, we set to map them using an array of cross-linking methods coupled with mass spectrometry, chemical mapping, and mutational analysis. We employed a long range cross-linker BS3 (bis(sulfosuccinimidyl)suberate, 11.4 Å), a medium range cross-linker DSG (disuccinimidyl glutarate, 7.7 Å), a short-range cross-linker Bpa (<4 Å) and a "zero-length" cross-linker EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide) (Leitner et al., 2010; Morozov et al., 2015).

Mass spectrometry analysis of the long-range cross-links obtained with BS3 and DSG revealed proximity of the inter-domain linker (residues K144, K153 and K156) of TEFM (Figure 30A, Figure 36A) to the specificity loop (residues K1087 and K1089) of mtRNAP - a structural element implicated in promoter specificity and formation of the RNA exit pore in single subunit RNAPs (Schwinghammer et al., 2013; Temiakov et al., 2000). We also detected a single interaction point between functionally important TEFM domains (CTD/linker) and mtRNAP (Figure 30A) using the EDC "zero-length" cross-linking, which relies on modification of D-K and E-K salt bridges (Rivera-Santiago et al., 2015). This interaction involves surface residue E214 (or E215) in TEFM and the exposed residue K1089 in the specificity loop of mtRNAP. To confirm these findings and to further map the interaction interface, we used the photo reactive amino acid p- benzoyl-phenylalanine (Bpa) incorporated into different positions within mtRNAP (Figure 30B, Figure 36B). Consistent with the EDC cross-linking data, we found that substitution of residues V1088 and I1091 in the specificity loop of mtRNAP resulted in efficient cross-linking with TEFM (Figure 30B). In addition, we detected a strong cross-link (~30%) to TEFM when Bpa was placed at position 610 in the intercalating hairpin of mtRNAP, in agreement with the structural data (Figure 30B and Figure 26C). Both the specificity loop and the intercalating loop are found in close proximity to each other in the EC, and although the tip of the specificity loop is not fully resolved in the EC-TEFM structure, it can be confidently placed in proximity to TEFM based upon RNA-mtRNAP cross-linking data (Figure 30D) (Schwinghammer et al., 2013).

To refine the interaction interface between mtRNAP and TEFM, we used a reciprocal approach and incorporated the Bpa probe into various positions in TEFM (Figure 30C, Figure 36B and C). We detected efficient cross-linking when Bpa was positioned in an unstructured, solvent exposed loop (the "docking loop", residues 305-315) at the distal end of the CTD dimer (Figure 30D). There, several TEFM Bpa variants (Bpa308, 309, 310, and 312) produced efficient

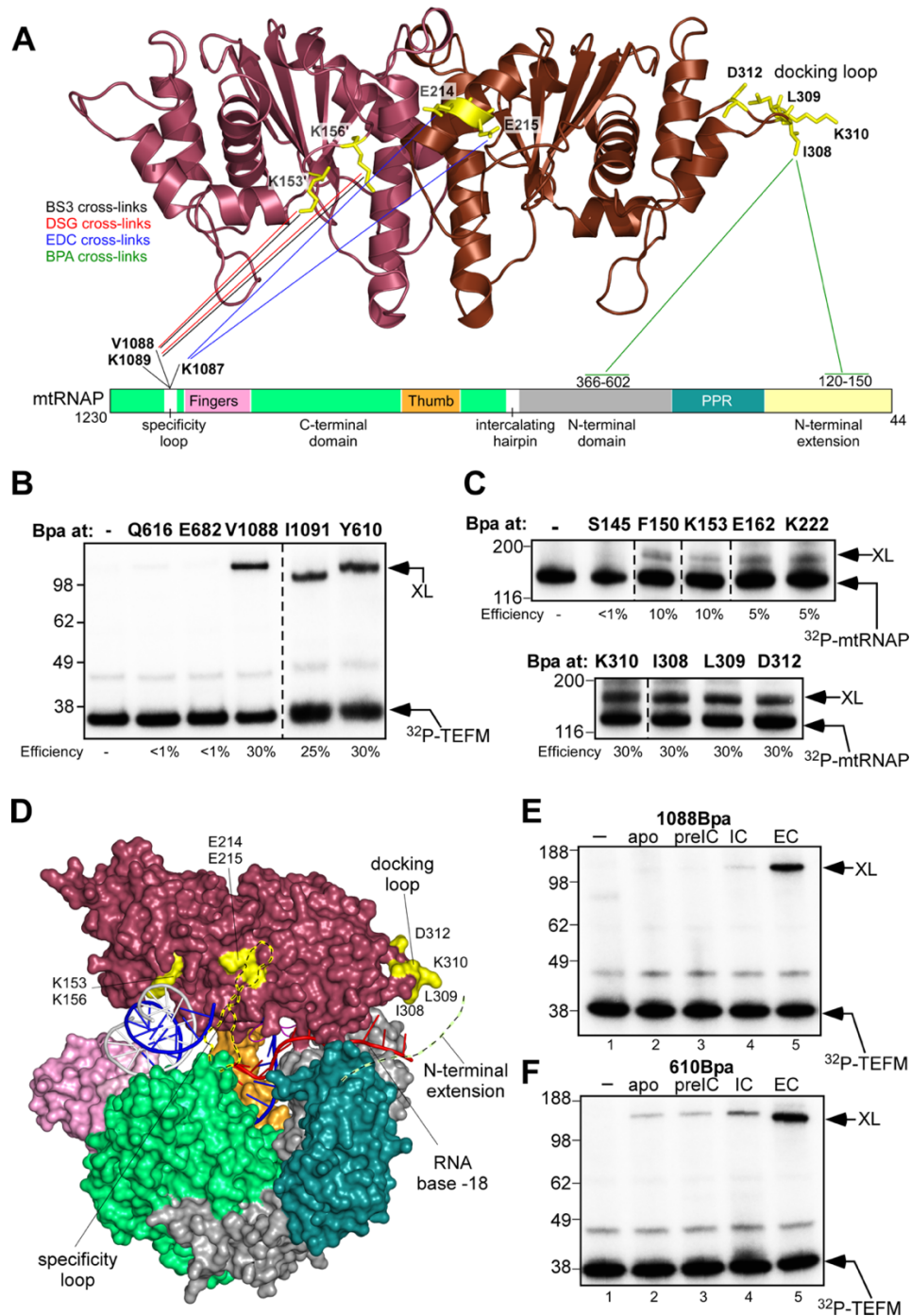


Figure 30. TEFM interactions revealed by cross-linking.

(A) Interaction regions in TEFM and mtRNAP revealed by chemical cross-linking and mass spectroscopy. The TEFM CTD is shown as a ribbon model; mtRNAP is depicted schematically as a bar with the major domains indicated as in Figure 26A. Cross-linked residues in TEFM are shown as sticks (yellow), lines represent the respective target residues/regions in mtRNAP. The colors of the lines indicate the cross-linking reagent used. (B) TEFM interacts with the specificity loop and the intercalating hairpin of mtRNAP. The ECs were assembled using mtRNAP containing Bpa at positions indicated and ^{32}P -labeled TEFM. The efficiency of cross-linking is indicated beneath the gel, size of molecular weight markers (kDa) is shown to the left of the gel. (C) Bpa cross-link reveals interaction of the docking loop of TEFM with mtRNAP. The ECs were assembled using ^{32}P -labeled mtRNAP and TEFM containing Bpa at positions indicated. (D) Location of the mapped regions on EC-TEFM structure. Cross-linking sites on TEFM (surface representation) are highlighted in yellow, the specificity loop (modeled; ribbon view) of mtRNAP is shown in yelloworange. The N-terminal extension domain of mtRNAP implicated in cross-linking with the docking loop of TEFM is shown as a dash line

(chartreuse). The RNA is extended by 9 nucleotides to indicate a probable RNA exit channel according to the RNA cross-linking data (Agaronyan et al., 2015; Schwinghammer et al., 2013). **(E,F)** TEFM can efficiently cross-link only to the EC. Cross-linking was performed using 32 P-labelled TEFM, 1088Bpa-mtRNAP (E) or 610Bpa-mtRNAP (F). The reaction mixtures contained mtRNAP (apo, lane 2), mtRNAP, TFAM and LSP (pre-initiation complex, lane 3), mtRNAP, TFAM, TFB2M and LSP (initiation complex, lane 4) or mtRNAP and RNA/DNA scaffold (EC, lane 5). The control reaction (lane 1) contained TEFM only.

(>30%) cross-linking to mtRNAP, suggesting an additional point of interaction between the two proteins (Figure 30A and C). We found that 308Bpa TEFM cross-links to two regions of mtRNAP that involve residues 120-150 in the N-terminal extension region of mtRNAP (65% of the cross-link) and residues 364-602, located in the palm subdomain, near to the PPR domain (35% of the cross-link) (Figure 37A to G).

Interestingly, the region 120-150 in mtRNAP (not present in the mtRNAP construct used to determine the EC-TEFM structure) constitutes the binding site for transcription initiation factor TFAM (Morozov et al., 2014), suggesting that binding of TEFM contributes to the transition from the initiation to the elongation phase of transcription. To demonstrate that TEFM binding is specific to the elongation phase of transcription, we assembled different transcription complexes - the pre-initiation complex, the initiation complex and the EC - using Bpa-modified mtRNAP variants (Figure 30E and F). Efficient cross-linking between mtRNAP and TEFM was only observed when the EC was formed (Figure 30E and F). In agreement with these data, analytical size exclusion chromatography demonstrated that TEFM does not efficiently bind to the initiation complex (Figure 31). To conclude, the cross-linking data above corroborate the EC-TEFM structure and expand the comprehensive map of interactions between TEFM and mtRNAP during transcription elongation.

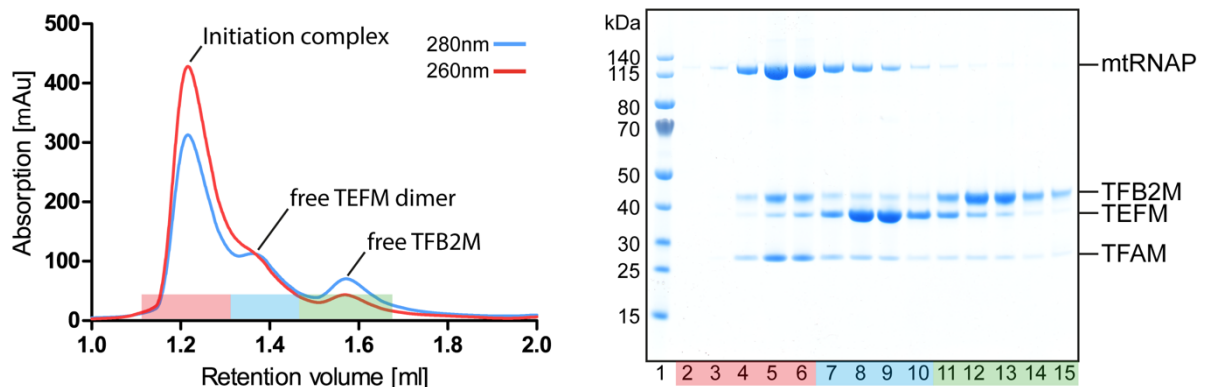


Figure 31. TEFM does not bind to the initiation complex.

Elution profile and SDS-PAGE of analytical size-exclusion chromatography of the initiation complex (LSP/mtRNAP/TFAM/TFB2M) in the presence of TEFM. Fractions analyzed by SDS-PAGE are indicated as colored bars in the chromatogram and below the SDS gel.

Mechanisms of transcription termination and anti-termination at CSBII

Human mtDNA encodes a G-rich sequence followed by an 8 bp "spacer" region and a 9 bp AU-rich region that precedes the termination point and constitutes the RNA-DNA hybrid in the EC (Figure 32A). CD spectroscopy experiments have confirmed that the G-rich region in RNA is sufficient to form a strong G-quadruplex structure (Wanrooij et al., 2012), which may fold into a parallel-stranded structure having three G-tetrad layers (Figure 32B) (Mukundan and Phan, 2013). The spacer region appears to be dispensable as it can be removed without affecting efficiency or site of termination (Figure 32C). However, when the AT-rich region in the CSBII sequence was changed to a GC-rich region, no termination was observed at the usual point, suggesting the importance of the weak RNA-DNA hybrid in transcription termination at CSBII (Figure 32D). Instead, multiple points of termination downstream of the usual termination site were observed, all of which corresponded to the occurrence of U-stretches in the RNA (Figure 32D). This indicates that efficient termination at CSBII depends on the formation of a G-quadruplex at the point at which it is most disruptive to the EC (i.e. in proximity to the RNA-DNA hybrid). Termination at the G-quadruplex in human mitochondria is therefore reminiscent of transcription termination in bacteria and T7-like phages, where secondary structure in the RNA (hairpin) is thought to destabilize the EC and a U-stretch is required for its disruption (Epshtein et al., 2007; Gusarov and Nudler, 1999; Ma et al., 2005).

The structure of the anti-termination complex combined with cross-linking and RNase footprinting data suggests that TEFM binds near the RNA exit channel and "guides" the emerging transcript from the region of its separation from the DNA to the point where RNA reaches a length of about 18-20 nucleotides (Figure 24, Figure 25, Figure 26, Figure 32). TEFM interacts with two mobile elements in mtRNAP - the intercalating hairpin and the specificity loop (Figure 30B and D), which were shown to contribute critically to EC stability (Schwinghammer et al., 2013). Modeling of an RNA G-quadruplex into the EC structure suggests that this bulky secondary structure would clash with both the specificity loop and the intercalating hairpin (Figure 32E). This would likely result in displacement of these elements and pulling of RNA from the active site, thus causing dissociation of mtRNAP from the DNA (Figure 32E). In contrast, binding of TEFM to the intercalating hairpin and specificity loop occludes the space in the EC where formation of the G-quadruplex would have the most dramatic impact on the complex stability (Figure 32F). Interestingly, mtRNAP does not efficiently terminate at hairpin termination signals, such as bacterial TrpA and T7 terminator T Φ (Figure 38), possibly due to the relatively small size of these hairpins (~23 Å in diameter vs. ~36 Å for the G-quadruplex), which may not be sufficient to disrupt mtRNAP elements at the point of termination (Figure 32E and F).

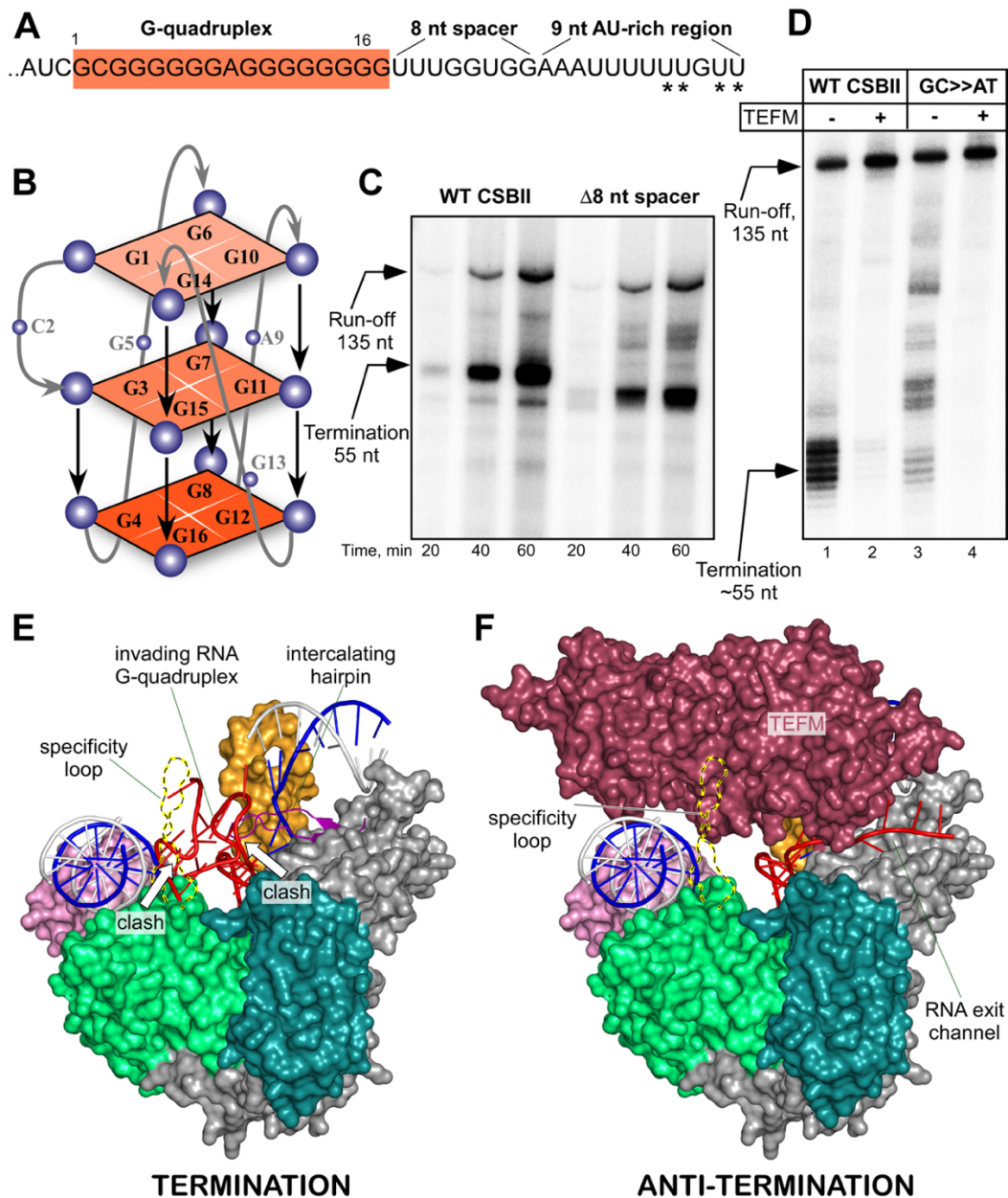


Figure 32. Mechanism of transcription termination at CSBII.

(A) Schematics of the RNA transcript encoded in the CSBII region in human mitochondria. The G-quadruplex region (G6AG8 variant) is highlighted in red, the points of termination are indicated by asterisks. (B) Schematic depiction of a possible RNA G-quadruplex structure having a single nucleotide bulge. (C) The spacer region of CSBII is dispensable for termination of transcription. Transcription assays were performed using promoter templates with the WT and mutated CSBII sequence, in which the spacer region was deleted. (D) The U-run encoded in the CSBII region is required for transcription termination. The transcription assay was performed using the WT and modified CSBII sequence (GC>>AT), in which the AT-rich sequence proximal to the termination site was substituted to the GC-rich sequence. (E) RNA G-quadruplex structure clashes with the specificity loop and intercalating hairpin of mtRNAP. The model is obtained by extending the RNA in mtRNAP EC with the 16 nucleotide-long G-quadruplex (PDB ID: 2M4P) (Mukundan and Phan, 2013). The G-quadruplex (red) is shown formed in vicinity of the trailing edge of transcription bubble in EC (surface representation). The specificity loop is modeled based on TEFM-mtRNAP Bpa cross-linking data. (F) Structure of the anti-termination complex (surface representation). Coloring as in Figure 26A.

3.2.3. Discussion

The bipartite role of TEFM in transcription elongation and anti-termination

The structure of the anti-termination complex explains how TEFM increases the processivity of mtRNAP. First, TEFM enhances interaction of the EC with the downstream DNA, contributing to a sliding clamp, a characteristic feature of a processive polymerase. Second, TEFM sequesters the single-stranded portion of the non-template DNA strand, thus preventing collapse of the transcription bubble. Third, because of the binding of TEFM to the specificity loop and the intercalating hairpin, which contributes to the formation of the RNA exit channel, the anti-termination complexes possess a much higher stability and processivity, as required for synthesis of genome-size RNA transcripts. Finally, TEFM may stabilize the post-translocation state of the EC, which is known to enhance RNAP processivity by promoting forward translocation and preventing pausing and arrest of transcription complexes on various sequences (Herbert et al., 2010).

The G-rich sequence found in CSBII is predicted to form a G-quadruplex structure and is likely the major cause of EC destabilization leading to termination of RNA synthesis and primer formation for mtDNA replication. Two hypotheses involving the nature of the G-quadruplex in human mitochondria have been advanced – that formation of the quadruplex involves a hybrid between the NT DNA strand and the RNA and, alternatively, that the quadruplex involves only RNA (Wanrooij et al., 2010; 2012). Analysis of the mtRNAP EC structure suggests that formation of the hybrid quadruplex structure is highly unlikely due to steric hindrance provided by mtRNAP, which separates the RNA from the DNA-RNA hybrid and diverts it away from the upstream DNA (Schwinghammer et al., 2013). Consistently, substitution of GMP to 7-deazaGMP (which disrupts G-quadruplex formation) in DNA does not affect termination at CSBII, suggesting that a RNA quadruplex is solely responsible for termination (Kuzmine et al., 2001; Wanrooij et al., 2012).

The efficiency of termination at CSBII has been demonstrated to depend upon the arrangement of the G residues in this region, which is highly polymorphic in the human mtDNA population. Thus, the more commonly occurring sequences - G6AG8 and G6AG7 (30% of the population) - exhibit significantly higher efficiency of termination as compared to the relatively rare G5AG8 and G5AG6 sequences (Kang et al., 2016; Tan et al., 2016). This suggests that the strength of the G-quadruplex is likely a primary contributor to the efficiency of transcription termination and thus to replication and maintenance of mtDNA copy number (Agaronyan et al., 2015). The more common G-quadruplex sequences in human mtDNA allow assembly similar to a three-layer G-quadruplex structure having a single nucleotide insertion (bulge) (Mukundan and Phan, 2013) (Figure 32B). This would result in a bulky structure forming inside the RNA exit channel of mtRNAP, likely causing disruption of the EC. Lack of termination of mtRNAP at hairpin terminators, at which a related T7 RNAP efficiently terminates, suggests that weaker secondary structures are tolerated by mtRNAP. This phenomenon may be attributed to the biological role of mtRNAP, which must transcribe through extensive secondary structures

of ribosomal and transfer RNAs. The structure of the anti-termination complex reveals that TEFM binds at the junction between the RNA-DNA hybrid and the upstream DNA, near the RNA exit channel (Figure 26A and B). This is predicted to prevent formation of the G-quadruplex structure near the point of RNA separation from the RNA-DNA hybrid, the mtRNAP region that defines EC stability. Formation of the G-quadruplex past this point would not impose as dramatic an effect on EC stability and, as consequence, cause efficient termination (Figure 32D). The anti-termination mechanism for TEFM is thus reminiscent to that of the phage lambda protein N and bacterial transcription factor RapA, both of which bind at the RNA exit channel and prevent RNA hairpin formation within the bacterial RNAP EC (Gusarov and Nudler, 2001; Liu et al., 2015).

TEFM as a dedicated transcription factor

Since the TEFM CTD shows high structural similarity to RuvC-like resolvases, the question as to whether it possesses a nuclease activity may have significant bearings on the mechanism of replication in human mitochondria. Indeed, the closest structural homologs of TEFM, *S.pombe* and *S.cerevisiae* Cce1, are functional resolvases (Doe et al., 2000; Lockshon et al., 1995). Yeast cells deficient in Cce1 exhibit a dramatic increase in mtDNA aggregation (Doe et al., 2000; Piskur, 1997). The crystal structure of the TEFM CTD dimer reveals two putative nuclease sites, which are similar in overall shape to the active sites of Cce1 (Figure 33A and B). *S.pombe* Cce1 uses three negatively charged catalytic residues (D320, E117 and D46) to coordinate Mg ions required for nuclease activity (Ceschini, 2001). In TEFM, the two corresponding residues, D342 and E232, are preserved; however, the third catalytic residue is altered to valine (V174), which is conserved in TEFM from all species (Figure 33A and B). The structure suggests that this substitution renders TEFM unable to coordinate a Mg ion. Indeed, no electron density compatible with a bound divalent cation was observed in our structure even though crystallization conditions included 200 mM MgCl₂. We therefore conclude that the nuclease sites in TEFM are degenerate and not functional. Indeed, considering that mammalian mitochondrial genomes are among the smallest, it is plausible that during evolution and transfer of many mitochondrial genes to the nucleus the resolvase function became non-essential for mtDNA maintenance. In contrast, preservation of a functional Holliday junction resolvase in yeast mitochondria is related to its essential function during replication, as the heritable mtDNA units in *S. cerevisiae* consist of genomes linked via recombination junctions (Doe et al., 2000; Lockshon et al., 1995). Although lacking nuclease activity, TEFM evolved to retain the ability of a Holliday junction resolvase to bind DNA, as the protein establishes simultaneous contacts with the RNA-DNA hybrid duplex and the downstream and upstream DNA regions of the transcription bubble (Figure 26A, B and D). These three duplexes of the EC are sharply bent relative to each other and resemble the topology of the DNA junctions, at which RuvC-like resolvases operate (Górecka et al., 2013).

Target specificity of TEFM

Many transcription factors can recognize a particular conformation of RNAP allowing for transient regulation of its activity (Washburn and Gottesman, 2015). This is clearly the case for TEFM, as the cross-linking data and analytical gel filtration assay suggest that TEFM can efficiently bind to the EC, but not to the pre-initiation complex, initiation complex or the apo mtRNAP (Figure 33C, Figure 30E and F). Our structural data provide an explanation for this observation. In the EC, the orientation of the two major structural elements of mtRNAP that bind TEFM - the intercalating hairpin and the specificity loop - are fixed by interactions with the DNA and RNA (Figure 26A and B). In contrast, in the apo mtRNAP structure, both loops are disordered (Ringel et al., 2011) and therefore cannot contribute to the efficient binding of TEFM (Figure 30E and F). In the preIC and the IC the intercalating hairpin and specificity loop are likely engaged in interactions with promoter DNA and may not be fully accessible to TEFM (Morozov et al., 2015). Once mtRNAP completes the transcription cycle and dissociates from mtDNA, TEFM can no longer bind to the mtRNAP efficiently, and dissociates prior to engagement of mtRNAP in transcription initiation (Figure 33C).

Our data on the mechanism of TEFM anti-termination activity, EC stabilization, mtRNAP processivity enhancement, and target specificity provide the molecular basis for the proposed role of TEFM in regulation of the switch between transcription and replication in human mitochondria (Agaronyan et al., 2015). Future experiments should be directed towards deciphering whether this regulation includes post-translational modifications of TEFM or changes in TEFM concentration in mitochondria by expression, degradation or other, yet unidentified factors.

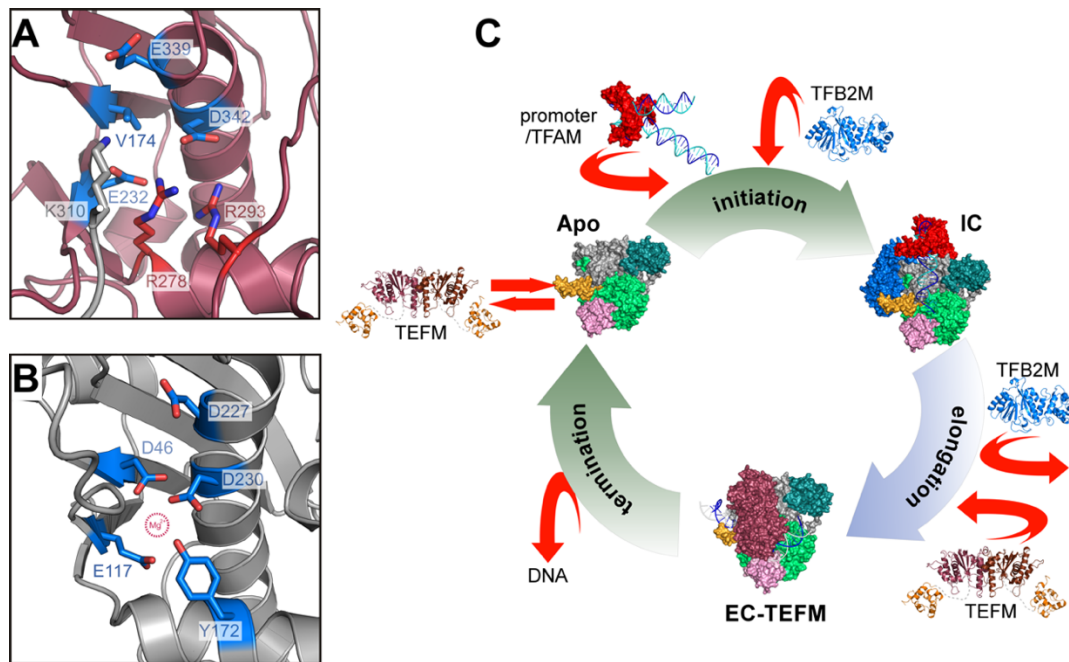


Figure 33. TEFM is a pseudo-resolvase.

(A,B) TEFM is a pseudo-resolvase. Putative active sites of human TEFM (A) and fission yeast Holliday junction resolvase Cce1 (B) are shown in similar orientation (ribbon representation). One of the magnesium-coordinating residues involved in catalysis in Cce1, D46, is replaced by the conserved residue V174 in TEFM. Two arginine residues conserved in all TEFM (R278 and R293, red sticks) are engaged in salt bridges with E232 and D342, effectively neutralizing their negative charge required for metal ion binding. An additional positive charge is contributed by K310 from the docking loop of a crystallographically related TEFM molecule (grey). (C) Schematics of the transcription cycle in human mitochondria. During initiation, mtRNA polymerase is recruited to promoter by interactions with TFAM. Subsequent binding of TFB2M results in formation of the initiation complex. Transition to the elongation phase of transcription occurs upon release of TFB2M and recruitment of TEFM, required for processive elongation. At the end of the transcription cycle, mtRNA polymerase dissociates from mtDNA, which triggers release of TEFM. Structures shown are of TFAM/LSP (PDB ID: 3TMM) (Ngo et al., 2011; Rubio-Cosials et al., 2011), the IC and the TFB2M models (Morozov et al., 2015), EC (PDB ID: 4BOC) (Schwinghammer et al., 2013), EC-TEFM (this work) and apo mtRNA polymerase (PDB ID: 3SPA) (Ringel et al., 2011).

4. Conclusion and Outlook

Expression and replication of the human mitochondrial genome depends on the mitochondrial transcription machinery. Thus, mitochondrial transcription is at the heart of cellular energy production and mitochondrial organelle maintenance. In multi-subunit RNAP transcription systems, initiation and elongation are the two most regulated steps, and this is likely to also be the case in mitochondria. However, since this system shares no homology with other factor-dependent transcription systems, the mechanistic interplay of the various factors with mtRNAP during initiation and elongation remained elusive, largely due to the lack of structural information.

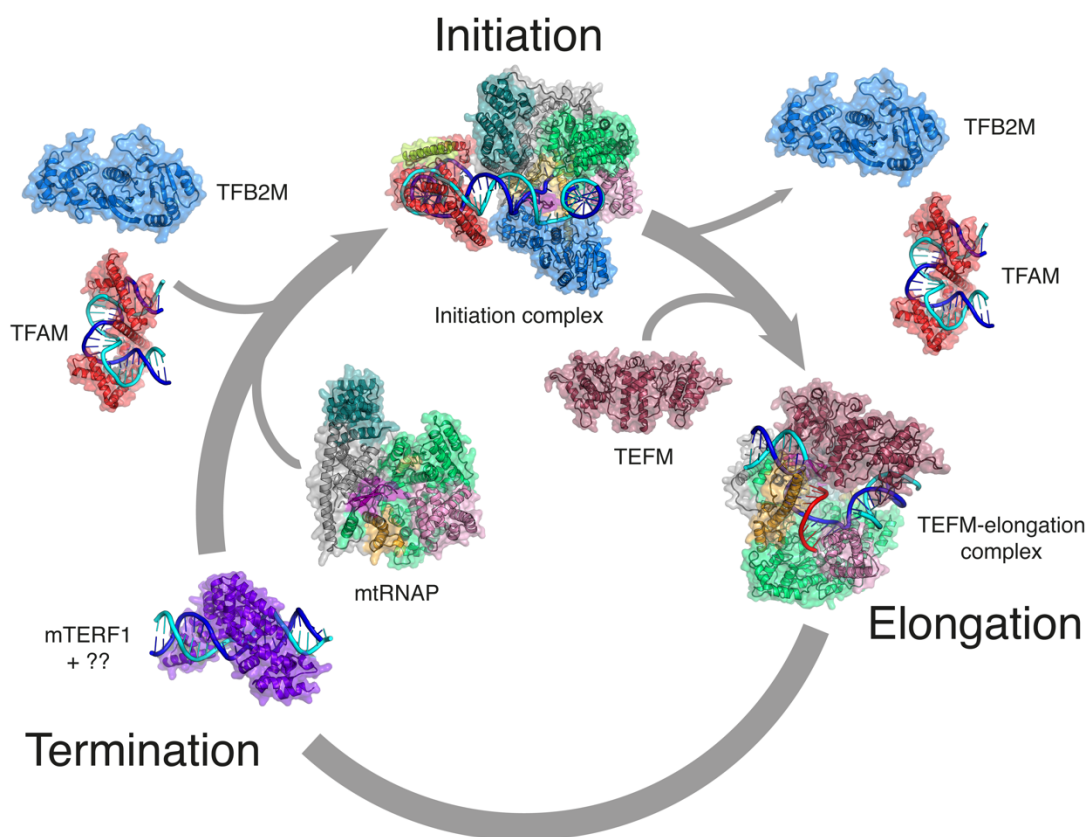


Figure 34. Structural view on the human mitochondrial transcription cycle.

Schematic depiction of the transcription cycle in human mitochondria. Structures are shown in ribbon representation with transparent surfaces. Note that structures are not shown to scale. Sources of structures used: TFAM: PDB ID: 3TMM (Ngo et al., 2011); mtRNAP: PDB ID: 3SPA (Ringel et al., 2011); mTERF1 PDB ID: 3MVA (Yakubovskaya et al., 2010); EC: PDB ID: 4BOC (Schwinghammer et al., 2013); TFB2M – this work; TEFM – this work; initiation complex – this work; TEFM-elongation complex – this work.

In this study, we extend our knowledge of human mitochondrial transcription by solving the structures of TFB2M and TEFM as well as of two important functional complexes, the initiation complex and the elongating anti-termination complex. Taken together, these data complement the previously elucidated structures and allow us to depict the transcription cycle in human mitochondria with all thus-far known players using structural snapshots of the individual steps

(Figure 34). Comparison of the structures presented here with their functional equivalents from other transcription systems illustrates that mitochondrial transcription is evolutionarily unique, but nonetheless utilizes comparable strategies as other systems to achieve the individual steps in transcription. This indicates that the mechanisms employed by transcriptional machineries and their regulation are dictated by the biophysical properties of nucleic acids and the biochemistry of the reaction catalyzed during RNA synthesis.

Taken together, these data represent major advances towards understanding the mechanistic basis underlying mitochondrial gene expression and lay the groundwork for future structural and biochemical studies aimed at fully understanding the mechanistic details and regulation of this important process.

4.1. Towards high-resolution structural data on the mitochondrial IC

The crystal structure of the human mitochondrial transcription initiation complex presented in this work unravels the architecture of the complex, explains how mtRNAP is tethered to the promoter by TFAM and suggests how TFB2M acts to stabilize the open promoter complex. Due to the limited resolution, however, some details remain elusive. In particular, we were not able to observe density for the entire melted template DNA strand. High-resolution structural data is necessary to visualize the interactions of the template DNA between positions -4 and +2 with the active site of mtRNAP. This will give further mechanistic insights, as it is likely that promoter recognition is in part achieved by interactions of mtRNAP with bases -2 and -3 in the template strand DNA (unpublished biochemical data by the Temiakov Lab).

A second interesting question that will require higher resolution structural data in order to be answered concerns the interaction between TFB2M and the single-stranded non-template DNA strand. While the data presented in this work clearly show that TFB2M interacts with this region of the DNA, details of the interaction remain invisible at the obtained resolution. Does TFB2M merely stabilize the single strand non-specifically by interacting with the phosphate backbone, or are there more specific interactions involved, such as base-flipping and stacking interactions? Notably, bacterial σ -factor, whose function seems to share conceptual similarities to TFB2M, recognizes specific bases (-35 and -11) within the non-template strand by base-flipping (Feklistov and Darst, 2011; Zhang et al., 2012).

To address these open questions, high-resolution structural data is needed. Although it was not possible to obtain such data during the course of this work, the structures presented here may guide future efforts towards obtaining such data. Two possible avenues to achieve this are imaginable, and may be combined if necessary.

Firstly, the structure of the IC may be used to guide future optimized crystallization trials. Limited diffraction of macromolecular crystals is usually caused by intrinsic disorder within the crystal lattice, and this in turn is often caused by flexible regions in the macromolecule or by imperfect packing within the crystal. In some cases, these parameters can be improved by optimizing the protein and DNA variants used. Although many potential modifications, both

pre- and post-crystallization, have been tested over the course of this work, no crystals displaying improved diffraction properties could be obtained. However, the newly obtained structure can be used to narrow down future attempts to obtain novel crystal forms. With the knowledge of how TFB2M interacts with mtRNAP and the DNA, one possibly promising direction would be to attempt crystallization of an mtRNAP-TFB2M-DNA complex, thus reducing size and complexity of the sample. An important result of the work presented here is the identification of an individually crystallizable variant of TFB2M. This paves the way for future structural experiments to obtain a more detailed picture of the interaction of TFB2M with nucleic acid. It seems particularly attractive to attempt co-crystallization of this variant with a single-stranded DNA molecule resembling the non-template strand in the IC, as this is likely to yield a more detailed picture of the involved interactions.

Secondly, another approach to obtain high-resolution structural data would be to employ a completely different method. Until recently, X-ray crystallography has dominated the structural biology field, as it was the only method capable producing structural data of large macromolecular assemblies at atomic resolution. This situation, however, has drastically changed over the last few years (Kühlbrandt, 2014). With the advent of direct detector technology and powerful new software, cryo-electron microscopy has advanced to the point where it can be used to routinely obtain structural data at near-atomic resolution. Thus, cryo-electron microscopic analysis of the mitochondrial initiation complex may be a more feasible avenue to obtain high-resolution structural data, as it circumvents the necessity to search for a crystal form with better diffraction properties. This approach bears the advantage that structural heterogeneity can be addressed *in silico* by sorting particles into different classes, ideally obtaining structural snapshots of different naturally occurring conformations.

4.2. A complete picture of mitochondrial transcription initiation

The structure of the open initiation complex represents a major milestone in our understanding of mitochondrial transcription. However, in order to gain a complete picture of this process, structural snapshots of two additional steps of initiation would be of particular interest.

Firstly, the closed pre-initiation complex consisting of mtRNAP, TFAM and promoter DNA would provide valuable insight into the mechanism of promoter binding. From the currently available data, it is not possible to infer what this complex looks like. In particular, it is unknown whether the marked distortion in the DNA between TFAM and mtRNAP is induced by binding of the polymerase already in the closed DNA state, or whether the observed trajectory is a hallmark of the open DNA. Modeling of a closed DNA duplex into the open promoter complex structure obtained in this study reveals a clash of the downstream DNA with the fingers domain in mtRNAP. This is not surprising, as the trajectory of the downstream duplex is dramatically shifted in the open complex. It is, however, not known whether this movement occurs spontaneously upon DNA opening or whether it is induced by the polymerase. Unfortunately, structure determination of this complex may prove difficult, as the interactions in the preIC may be of very transient nature, making the complex inherently unstable. Indeed, experiments

conducted during this thesis work indicate that both the preIC as well as the IC are significantly less stable when formed on a fully duplexed DNA, as compared to the artificial mismatch bubble used for structure determination.

The second important snapshot missing is that of an initially transcribing complex (ITC), consisting of mtRNAP, TFAM, TFB2M, promoter DNA and a short piece of RNA. Such complexes have been isolated and characterized structurally for other transcription systems (Cheetham and Steitz, 1999; Cheung et al., 2011; Engel et al., 2017) and provide important mechanistic insights into *de novo* RNA synthesis. Comparison of the open IC structure and the EC-TEFM structure obtained in this work demonstrates a profound structural rearrangement of the upstream DNA during promoter escape. However, the mechanism of how this rearrangement is induced remains speculative, and it would therefore be highly instructive to obtain structural information of the IC with different lengths of initial RNA chains synthesized. A structure of an ITC or an open complex bound to an incoming nucleotide is also likely to expand our knowledge on the functions of TFB2M, as its N-terminus has been cross-linked to the priming nucleotide (Sologub et al., 2009). This region did not show clear density in the open IC structure, indicating that it is mobile in the absence of the priming nucleotide. Thus, formation of a complex with a non-hydrolyzable nucleotide analogue or a short initial RNA chain may stabilize the TFB2M N-terminus and make it amenable to structural analysis.

Taken together, such future studies aimed at obtaining structural snapshots of each step of transcription initiation hold promise to ultimately allow generation of a comprehensive “movie” of the events leading to RNA synthesis.

4.3. Further studies on TEFM function and regulation

The structures of TEFM and the EC-TEFM complex presented in this study clarify how this factor enhances transcriptional processivity of mtRNAP and confers resistance against premature termination at CSBII. It remains an open question, however, what the precise role of the N-terminal domain of TEFM is, as it is dispensable for the transcriptional functions studied here. It is possible that other regions in the mitochondrial genome exist that were not explicitly analyzed in this study, but which require the N-terminal domain of TEFM for regulated transcription. Another possibility is that this domain is involved in an additional, thus far uncharacterized function of TEFM.

Another important aspect not yet understood is how TEFM is recruited to the polymerase, and how its activity is regulated. The biochemical and cross-linking data indicates that TEFM binds to mtRNAP following promoter escape and initiation factor clearance, but the precise sequence of events is not known. Clearly, the switch between transcription and replication in mitochondria must be regulated, and the most obvious target for regulation would be TEFM itself.

Therefore, future research should be focused on determining whether the activity of TEFM is dependent on post-translational modifications or whether its concentration in mitochondria is

regulated according to cellular needs. In addition, it should be investigated whether TEFM has additional, not yet recognized functions and interaction partners *in vivo*.

4.4. Evolutionary aspects of mitochondrial transcription

The fact that mitochondria utilize a bacteriophage-like RNAP instead of a multi-subunit RNAP related to eubacterial RNAPs is both surprising and puzzling. The evolutionary origin of this enzyme has not been conclusively established (Cermakian et al., 1997), but it is noteworthy that not only the mitochondrial RNA polymerase but also the replicative helicase and the DNA polymerase seem to be of bacteriophage origin (Shutt and Gray, 2006) and almost all eukaryotes utilize this machinery in their mitochondria. There are, however, notable exceptions, such as the freshwater flagellate *Reclinomonas americana*, whose mitochondrial genome is among the highest in coding gene content and contains several genes homologous to eubacterial multi-subunit polymerase genes (Lang et al., 1997). Although it is not known whether their gene products serve as a functional RNA polymerase, this finding raises interesting questions on the evolutionary origin of the bacteriophage-like mitochondrial transcription and replication systems.

The structural data obtained in this study highlight a further evolutionarily interesting aspect, as they provide two examples of functionally and evolutionarily unrelated protein folds being utilized for a function in mitochondrial transcription. The structure of TFB2M confirms its relationship to SAM-dependent rRNA methyltransferases and the structure of the initiation complex demonstrates how conserved properties of this fold, such as single-stranded oligonucleotide binding via a positively charged patch, are elegantly utilized for a function in transcription initiation. The fold of the transcriptionally active C-terminal part of TEFM closely resembles that of Holliday Junction resolvases, yet no resolvase or nuclease activity could be detected by us and others (Minczuk et al., 2011). Instead, the property of Holliday Junction resolvase domains to bind cruciform duplex-single strand nucleic acid junctions is apparently repurposed to yield a processivity factor for mitochondrial transcription. Interestingly, a similar situation may have arisen for the metazoan mitochondrial DNA polymerase γ , which associates with the accessory factor β for processive DNA replication. DNA polymerase β shows sequence- and structural homology to class IIa aminoacyl tRNA synthetases (Fan et al., 1999; 2006).

4.5. Clinical relevance of mtRNAP structures

In addition to the insights gained about TEFM function, the structure of the EC-TEFM complex further complements our previous knowledge of the mechanism of mitochondrial transcription elongation. In the previously solved EC structure, mtRNAP is in the pre-translocated state (Schwinghammer et al., 2013). The EC-TEFM complex structure now reveals the post-translocated state of mtRNAP. This is an important finding, as it demonstrates that the catalytic mechanism of RNA chain elongation is indeed conserved between bacteriophage and mitochondrial RNAPs. In the case of T7 RNAP, the availability of post-translocated elongation

complex crystals proved to be a prerequisite to obtaining structures of the polymerase with incoming substrate NTPs or product complexes (Tahirov et al., 2002; Temiakov et al., 2004; Yin and Steitz, 2002; 2004). Thus, crystallization of mtRNAP in this state paves the way for structural investigations of the interactions between the polymerase and incoming substrate nucleotides.

The next step will therefore be to attempt co-crystallization of the EC-TEFM complex with non-hydrolyzable substrate analogues, such as α,β -methylene-ATP, in order to obtain a structural snapshot of the molecular interactions that govern substrate selection. This may be of broad interest for clinical applications, as recent studies suggest that some of the severe side effects observed in clinical trials of antiviral ribonucleosides, which inhibit viral RNA polymerases, are due to off-target effects on mitochondrial RNA polymerase (Arnold et al., 2012; Feng et al., 2016). Hence, structure-guided drug design may prove powerful for the development of small molecule compounds capable of selectively inhibiting only viral polymerases, ultimately leading to new treatments for human diseases such as Hepatitis C.

5. Appendix

5.1. Extended Materials and Methods for the initiation study

5.1.1. Extended Methods for the initiation study

This section contains methods specifically related to the study on the structure of the human mitochondrial transcription initiation complex, which were performed by our collaborators in the Temiakov lab. Parts of the methods described here are currently in the process of peer review for publication:

H.S. Hillen, Y.I. Morozov, A. Sarfallah, D. Temiakov and P. Cramer (2017) Structural basis of mitochondrial transcription initiation. *Cell*, in revision

A detailed list of author contributions can be found on page VI.

Transcription assays

Standard transcription reactions were carried out using PCR-amplified DNA templates containing the LSP promoter (region -60 to +20) as described previously (Morozov et al., 2015). The reactions contained DNA templates (50 nM), $\Delta 119$ mtRNAP (50 nM), Cys-less TFAM (50 nM), $\Delta 20$ TFB2M (50 nM) in a transcription buffer containing 40 mM Tris-HCl pH 7.9, 10 mM MgCl₂ and 10 mM DTT in the presence of ATP (0.3 mM), GTP (0.3 mM), UTP (0.01 mM) and 0.3 μ Ci [γ -³²P] UTP (800 Ci/mmol). To assay the activity of the IC assembled on pre-melted LSP, the reaction was performed in the presence of ATP (0.3 mM), GTP (0.3 mM) and 0.3 μ Ci [γ -³²P] ATP (800 Ci/mmol) to generate 4-5 nt RNA products. Reactions were carried out at 35 °C for 30 min and stopped by addition of an equal volume of 95% formamide/0.05 M EDTA. The products were resolved by 20% PAGE containing 6 M urea and visualized by PhosphorImager (GE Healthcare).

5.1.2. Extended figures and tables of the initiation study

This section contains tables related to the X-ray crystallographic analysis, which was performed as part of this thesis. The data in this section are currently under peer review for publication:

H.S. Hillen, Y.I. Morozov, A. Sarfallah, D. Temiakov and P. Cramer (2017) Structural basis of mitochondrial transcription initiation. *Cell*, in revision

A detailed list of author contributions can be found on page VI.

Table 12. X-Ray data collection and refinement statistics for TFB2M^{cryst} structure.

TFB2M ^{cryst}	
Data collection	
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	44.0, 165.7, 44.7
α , β , γ (°)	90.0, 98.0, 90.0
Resolution (Å) ^a	50 – 1.75 (1.79 – 1.75) ^a
<i>R</i> _{meas}	0.048 (2.874)
<i>I</i> /σ(<i>I</i>)	20.17 (0.75)
<i>CC</i> _{1/2}	100.0 (37.7)
Completeness (%)	97.8 (97.5)
Redundancy	7.1 (7.0)
Refinement	
Resolution (Å)	42.1 – 1.75
No. reflections	62087
<i>R</i> _{work} / <i>R</i> _{free}	0.175 / 0.209
No. atoms	
Protein	4627
Ligand/ion (Cl ⁻ , Glycerol)	8
Water	411
<i>B</i> factors (Å ²)	
Protein	52.24
Ligand/ion	85.22
Water	63.26
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.84
Ramachandran	
Preferred/allowed/ disallowed (%)	97.50 / 2.14 / 0.36

^a Values in parentheses are for the highest-resolution shell.

Table 13. X-Ray data collection, phasing and refinement statistics for IC LSP structure

	IC LSP Native	IC LSP SeMet all proteins ^b	IC LSP SeMet mtRNAP	IC LSP SeMet TFAM
Data collection				
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	103.9, 197.0, 137.2	103.9, 198.3, 136.1	103.8, 198.8, 138.0	103.6, 197.1, 136.6
α , β , γ (°)	90, 99.87, 90	90.0, 99.9, 90.0	90.0, 100.0, 90.0	90.0, 99.8, 90.0
Wavelength	0.99998	0.97856	0.97864	0.97856
Resolution (Å) ^a	50 – 4.50 (4.61 – 4.50) ^a	50 – 4.99 (5.11 – 4.99)	50 – 5.60 (5.74 – 5.60)	50 – 5.50 (5.63 – 5.50)
<i>R</i> _{meas}	0.061 (1.857)	0.167 (5.241)	0.094 (2.590)	0.086 (2.490)
<i>I</i> / σ (<i>I</i>)	14.46 (1.33)	19.94 (1.14)	14.53 (1.13)	24.71 (1.58)
<i>CC</i> _{1/2}	99.8 (54.0)	99.9 (51.3)	100.0 (55.2)	100 (72.4)
Completeness (%)	98.7 (96.1)	99.8 (99.2)	99.0 (99.3)	99.3 (99.0)
Redundancy	7.1 (7.0)	63.4 (56.0)	11.6 (11.9)	28.6 (28.6)
Refinement				
Resolution (Å)	49.59 – 4.50			
No. reflections	31982			
<i>R</i> _{work} / <i>R</i> _{free}	0.273 / 0.309			
No. atoms				
Macromolecules	28,039			
Ligand/ion	-			
Water	-			
<i>B</i> factors (Å ²)				
Macromolecules	350.1			
Ligand/ion	-			
Water	-			
r.m.s deviations				
Bond lengths (Å)	0.004			
Bond angles (°)	0.92			
Ramachandran				
Preferred/allowed/ disallowed (%)	95.86 / 3.40 / 0.73			

^a Values in parentheses are for the highest-resolution shell.^b Diffraction data from two crystals were merged.

Table 14. X-Ray data collection statistics for IC LSP Bromine derivatives.

	IC LSP Bromine 1^b	IC LSP Bromine 2^b	IC LSP Bromine 3	IC LSP Bromine 4
Data collection				
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	103.6, 196.7, 137.1	103.9, 195.6, 137.1	103.5, 196.4, 137.1	103.9, 197.7, 137.7
α , β , γ (°)	90.0, 99.7, 90.0	90.0, 99.8, 90.0	90.0, 99.7, 90.0	90.0, 99.6, 90.0
Wavelength	0.91930	0.91930	0.91930	0.91882
Resolution (Å) ^a	50 – 6.00 (6.15 – 6.00) ^a	50 – 6.00 (6.15 – 6.00)	50 – 6.00 (6.15 – 6.00)	50 – 6.01 (6.15 – 6.01)
<i>R</i> _{meas}	0.154 (1.401)	0.094 (1.010)	0.073 (1.137)	0.115 (3.226)
<i>I</i> / σ (<i>I</i>)	11.84 (1.38)	16.47 (2.52)	10.29 (1.25)	15.27 (0.84)
<i>CC</i> _{1/2}	99.9 (73.2)	100.0 (84.1)	99.9 (65.6)	100.0 (51.9)
Completeness (%)	99.7 (99.5)	99.7 (99.7)	98.7 (99.5)	99.7 (99.2)
Redundancy	12.8 (10.9)	12.5 (13.4)	3.6 (3.7)	21.2 (21.0)
	IC LSP Bromine 5	IC LSP Bromine 6		
Data collection				
Space group	P2 ₁	P2 ₁		
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	104.3, 196.7, 137.7	102.7, 196.4, 137.3		
α , β , γ (°)	90.0, 99.8, 90.0	90.0, 99.8, 90.0		
Wavelength	0.91880	0.91880		
Resolution (Å) ^a	50 – 6.01 (6.16 – 6.01)	50 – 7.00 (7.17 – 7.00)		
<i>R</i> _{meas}	0.08 (1.86)	0.09 (1.89)		
<i>I</i> / σ (<i>I</i>)	23.05 (1.63)	21.22 (1.83)		
<i>CC</i> _{1/2}	100.0 (79.2)	100 (79.4)		
Completeness (%)	99.6 (98.0)	99.7 (99.8)		
Redundancy	21.1 (20.5)	21.0 (22.0)		

^a Values in parentheses are for the highest-resolution shell.^b Diffraction data from two crystals were merged.

Table 15. X-Ray data collection, phasing and refinement statistics for IC HSP structure.

	IC HSP Native	IC HSP SeMet TFAM
Data collection		
Space group	P2 ₁	P2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	103.6, 197.7, 134.3	103.9, 199.4, 134.6
α , β , γ (°)	90, 99.39, 90	90, 99.57, 90
Wavelength	0.97864	0.97895
Resolution (Å) ^a	50 – 4.50 (4.61 – 4.50) ^a	50 – 6.50 (6.66 – 6.50)
<i>R</i> _{meas}	0.087 (2.232)	0.152 (2.146)
<i>I</i> / σ (<i>I</i>)	13.38 (1.11)	15.67 (1.58)
<i>CC</i> _{1/2}	100.0 (50.2)	99.9 (67.3)
Completeness (%)	99.1 (99.6)	99.7 (100.4)
Redundancy	7.0 (7.3)	21.3 (22.0)
Refinement		
Resolution (Å)	49.49 – 4.50	
No. reflections	31416	
<i>R</i> _{work} / <i>R</i> _{free}	0.289 / 0.331	
No. atoms		
Macromolecules	28045	
Ligand/ion	-	
Water	-	
<i>B</i> factors (Å ²)		
Macromolecules	322.63	
Ligand/ion	-	
Water	-	
r.m.s deviations		
Bond lengths (Å)	0.004	
Bond angles (°)	0.91	
Ramachandran		
Preferred/allowed/ disallowed (%)	95.60 / 3.54 / 0.87	

^a Values in parentheses are for the highest-resolution shell.

5.2. Extended Materials and Methods for the elongation study

5.2.1. Extended Methods for the elongation study

This section contains methods specifically related to the study on the mechanism of transcription anti-termination in human mitochondria and carried out predominantly by the collaborators from the Temiakov lab. Parts of the methods described here have been accepted for publication:

H.S. Hillen, A.V. Parshin, K. Agaronyan, Y.I. Morozov, J.J. Graber, A. Chernev, K. Schwinghammer, H. Urlaub, M. Anikin, P. Cramer and D. Temiakov (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell*, in press

A detailed list of author contributions can be found on page VI.

Limited proteolysis of TEFM and analysis of the cleavage products

TEFM was subjected to limited proteolysis using trypsin (Sigma, protease/TEFM ratio 1:20,000 w/w), ArgC protease (Sigma, 1:1000 w/w), LysC protease (Sigma, 1:1000 w/w) for 1 h at RT. Products of protease digestion of TEFM were dissolved in a solution containing 5% acetonitrile and 0.1% tetrafluoroacetic acid (TFA) and purified using C4 ZipTip. The peptide mixtures were analyzed using a Bruker MicroFlex MALDI-TOF instrument in sinapinic acid matrix with R=5000.

Crystallization of human TEFM N-terminal domain

TEFM NTD (10 mg/ml) was crystallized using the sitting drop vapor diffusion method and Crystallization Screen solution 15 (Hampton Research) containing 30% PEG4000, 0.2 M ammonium acetate, 0.1 M sodium acetate pH 4.0 to produce large (0.05 mm x 0.05 mm x 0.8 mm) needle-like crystals. The crystals were flash-frozen using stabilization solution containing 30% PEG4000, 15% PEG400, 0.1 M ammonium acetate and 0.05 M sodium acetate pH 4.0.

Promoter templates for transcription assays

Templates for transcription anti-termination assays containing the LSP promoter and CSBII region were prepared by PCR amplification as described previously (Agaronyan et al., 2015). For the processivity assay, plasmid DNA (pT7blue) containing the LSP promoter region (-70/+70LSP_pT7blue) (Morozov et al., 2015) was linearized using the XhoI nuclease. The template was purified using phenol-chloroform extraction followed by ethanol precipitation. For termination assays, T7 phage terminator T Φ and bacterial terminator TrpA were cloned

into LSP-containing plasmid (-70/+70LSP_pT7blue) 70 bp downstream of transcription start site. For the halted complex assays and RNA foot- printing, a variant of the LSP-promoter template (“+35LSP”) with the initial transcribed sequence AAAGATAAAATTTGAAATGGTAGTTGTTTAAGTTGC was used.

Transcription assays

Standard transcription reactions were carried out using PCR-amplified DNA templates (50 nM), mtRNAP (50 nM), TFAM (50 nM), TFB2M (50 nM) in a transcription buffer containing 40 mM Tris-HCl pH 7.9, 10 mM MgCl₂ and 10 mM DTT in the presence of ATP (0.3 mM), GTP (0.3 mM), UTP (0.01 mM) and 0.3 μ Ci[α -³²P] UTP (800 Ci/mmol). Reactions were carried out at 35 °C for 60 min and stopped by addition of an equal volume of 95% formamide/0.05 M EDTA. The products were resolved by 20% PAGE containing 6 M urea and visualized by PhosphorImager (GE Healthcare). In the processivity assay, the reactions (10 μ l) were stopped by addition of 190 μ l of stop buffer containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM EDTA, 1% SDS and 20 μ g/ml proteinase K. After incubation at 55 °C for 1 h the RNA products were extracted with phenol/chlorophorm/isoamyl alcohol mixture followed by ethanol precipitation and resolved by 6% PAGE containing 6 M urea.

Nuclease foot-printing assays

For RNA foot-printing, ECs were halted 35 bp downstream of the promoter start site by omitting CTP from the transcription reaction and treated with RNase 1f (New England Biolabs) for 5 min at RT. For DNA foot-printing, the ECs were assembled using R14/TS36/NT36 scaffold, incubated for 14 min with TEFM at RT and treated with DNase I (New England Biolabs) for 5 min at RT. The sequences of the oligos (IDT DNA) are as follows (5' to 3'): AGUCUGCGGCGCGC (RNA14), GGGCTTAGTTCGTCTGGCGTGCGCGCCGCTACACCATGTTTGCTGACC (TS26), GGTCAGCAAACATGGTGTAAGTATTACGACGCCAGACGAACCTAAGCCC (NT26A). The reactions were stopped by addition of an equal volume of 95% formamide/0.05 M EDTA and resolved by 20% PAGE containing 6 M urea.

Cross-linking using artificial photo reactive amino acid (Bpa)

The ECs (0.5 μ M) were assembled using R14/TS2/NT2 scaffold (Schwinghammer et al., 2013) and mtRNAP. TEFM (1 μ M) was added to the EC and incubated for 15 min at RT. The cross-linking was activated by UV irradiation at 312 nm for 30 min at RT.

Protein-DNA photo cross-linking

The cross-linking was performed as described in (Agaronyan et al., 2015) using a tag-less variant of mtRNAP (Δ 108 mtRNAP, (Morozov et al., 2014)) and RNA-DNA scaffold, containing a ³²P-labeled DNA strand with a 4-thio UMP probe.

Mapping of TEFM-mtRNAP interaction using disuccinimidyl glutarate (DSG) cross-linking

To map the TEFM binding site, EC-TEFM complex (5 μ M in 40 mM Na-HEPES pH 7.5) was treated with a mixture of DSG/[2 H $_4$]-DSG (0.6 mM DMSO solution, ProteoChem) and the products of the reaction separated using 7% Tris-glycine SDS PAGE. The cross-linked species were excised from the gel and subjected to trypsin digestion in 100 mM NH $_4$ HCO $_3$ buffer overnight at 37 °C. The products of the digestion reaction were extracted from the gel with CH $_3$ CN-0.1% TFA in water (9:1), dried in vacuum, re-dissolved in 10 mg/ml CHCA matrix in CH $_3$ CN-0.1% TFA in water (1:1), and the resulting solution applied on target for MALDI-TOF mass spectrometry. Mass spectra were taken on a MicroFlex LRF spectrometer (Bruker). Positively charged ions (M+H $^+$) were analyzed in the reflector mode (m/z 1000 to 4000) and cross-linked peptides identified using xBobcat mass matching search engine (http://prottools.ethz.ch/orinner/public/htdocs/xquest/index_review.html).

Mapping of the Bpa-TEFM cross-linking sites in mtRNAP

For mapping with the HRV 3C protease (Pierce), the ECs were assembled using Δ 119 mtRNAPs with the engineered 3C cleavage sites (LEVLFQGP) - 3C-PPR mtRNAP (substitution of residues 358-364, cleaved at Q363), 3C-thumb mtRNAP (insertion after residue H745, cleaved at Q751) and 3C-lever-loop mtRNAP (insertion after residue K596, cleaved at Q602). All these mtRNAP variants also contained an engineered protein kinase (PKA) site at the N-terminus (MGHHHHHHRRASV...) (Morozov et al., 2015). Upon UV irradiation, the ECs (0.5 μ M) were incubated with 3C protease (1:1 w/w ratio) for 15-25 h at 40 °C and the products of the reaction resolved using 4-12% SDS PAGE (Invitrogen). NTCB mapping of the TEFM sites in mtRNAP was performed as described in (Morozov et al., 2015). The TEFM-mtRNAP cross-link was excised from the gel, eluted with 0.2% SDS for 1 h at 4 °C and precipitated with 80% acetone. The dried material was dissolved in solution containing 0.2% SDS prior to the NTCB treatment.

Pyrophosphorolytic assay

The ECs (250 μ M) were assembled using RNA10/TS1/NT1 scaffold (RNA10: CUGCGCGCAU; TS1: GGGTCCTGTCTGAAATCGACATCGCCGC; NT1: CGATTTTCAGACAGGACCC) for 10 min at room temperature in the presence or absence of TEFM (500 μ M). The reactions were incubated with pyrophosphate (50-500 μ M) for 30 min at room temperature and stopped by the addition of formamide/EDTA containing buffer. The samples were resolved by electrophoresis in 20% PAGE in the presence of 6 M urea and visualized by PhosphorImager.

5.2.2. Extended figures and tables of the elongation study

This section contains extended figures related to the study on the mechanism of transcription anti-termination in human mitochondria, which were generated by our collaborators in the Temiakov lab. In addition, it contains tables related to the X-Ray crystallographic analysis performed as part of this thesis. Parts of the methods described here have been accepted for publication:

H.S. Hillen, A.V. Parshin, K. Agaronyan, Y.I. Morozov, J.J. Graber, A. Chernev, K. Schwinghammer, H. Urlaub, M. Anikin, P. Cramer and D. Temiakov (2017) Mechanism of transcription anti-termination in human mitochondria. *Cell*, in press

A detailed list of author contributions can be found on page VI.

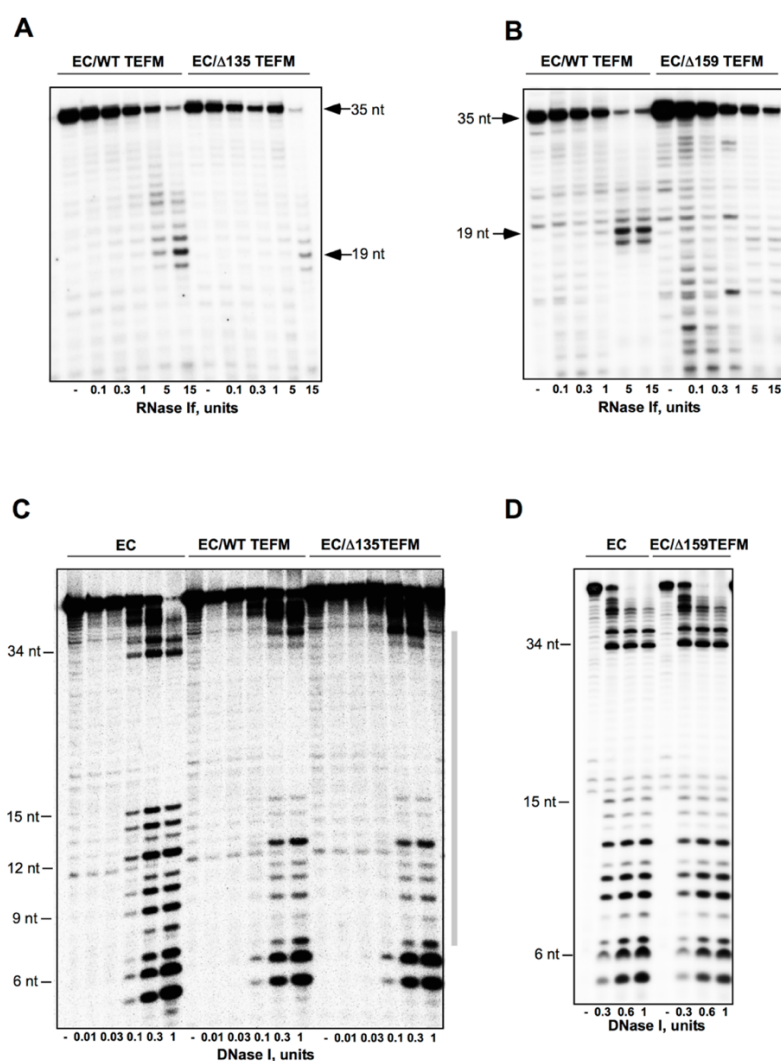


Figure 35. TEFM interacts with RNA and DNA in the EC.

(A) TEFM lacking the NTD domain protects 19-20 nt of RNA from RNase 1 cleavage in the anti-termination complex. The ECs were halted 35 bp downstream of the LSP promoter in the presence of

WT or $\Delta 135$ TEFM and treated with RNase 1 for 5 min at RT. **(B)** TEFM lacking the NTD domain and the inter-domain linker does not protect RNA from RNase 1 cleavage. The experimental conditions were as described above. **(C)** TEFM lacking the NTD domain protects the downstream and the upstream DNA in the EC. The EC was assembled using R14/TS26/NT26 scaffold in the absence of TEFM or in the presence of WT or $\Delta 135$ TEFM prior to DNase I treatment for 5 min at RT. **(D)** TEFM lacking the NTD domain and the inter-domain linker leaves no clear footprint in the DNase I protection assay. The experimental conditions were as described above.

A

Crosslinker	Protein 1	Protein 2	Residue 1	Residue 2	peptide	times detected
BS3	mtRNAP	TEFM	755	144	KAE LR(1)-KSPENR(1)	4
BS3	mtRNAP	TEFM	1089	144	VKQIGGGIQSITYTHNGDISR(2)-KSPENR(1)	6
BS3	mtRNAP	TEFM	1089	153	VKQIGGGIQSITYTHNGDISR(2)-KLLKPDIER(1)	3
BS3	mtRNAP	TEFM	1089	144	VKQIGGGIQSITYTHNGDISR(2)-LLKPDIER(3)	6
EDC	mtRNAP	TEFM	632	42	NVQQIGILKPHPAYVQLLEK(19)-KITPNVTFCDENAKEPENALDK(1)	6
EDC	mtRNAP	TEFM	632	42	NVQQIGILKPHPAYVQLLEK(19)-KITPNVTFCDENAK(1)	6
EDC	mtRNAP	TEFM	755	56	KAE LR(1)-EPENALDK(1)	7
EDC	mtRNAP	TEFM	1089	84	VKQIGGGIQSITYTHNGDISR(2)-ELEAFR(1)	3
EDC	mtRNAP	TEFM	1089	101	VKQIGGGIQSITYTHNGDISR(2)-SINIVEHR(6)	3
EDC	mtRNAP	TEFM	755	101	KAE LR(1)-SINIVEHR(6)	21
EDC	mtRNAP	TEFM	1206	124	ETLQAVPKPGAFDLEQVK(1)-YKSTVQVCNSILCPK(2)	30
EDC	mtRNAP	TEFM	1089	214	VKQIGGGIQSITYTHNGDISR(2)-GIYSSSVYLEEISSISK(10)	7
EDC	mtRNAP	TEFM	1089	215	VKQIGGGIQSITYTHNGDISR(2)-GIYSSSVYLEEISSISK(11)	22
DSG	mtRNAP	TEFM	1087	153	LDSKVK(4)-KLLKPDIER(1)	N/A
DSG	mtRNAP	TEFM	1087	156	LDSKVK(4)-KLLKPDIER(4)	N/A

B**Bpa-mtRNAP**

mtRNAP	cross-link efficiency
T421Bpa	<1%
T433Bpa	<1%
C591Bpa	<1%
R601Bpa	<1%
Y607Bpa	<1%
Y610Bpa	30%
Q616Bpa	<1%
Q617Bpa	<2%
E682Bpa	<1%
Q728Bpa	<1%
R770Bpa	<1%
S774Bpa	<1%
E778Bpa	<1%
E889Bpa	<1%
E890Bpa	<1%
K988Bpa	<1%
V1088Bpa	30%
I1091Bpa	25%
K1189Bpa	<1%

BpaTEFM

TEFM	cross-link efficiency
R89Bpa	<1%
K137Bpa	<1%
R140Bpa	<1%
K142Bpa	<1%
S145Bpa	<1%
F150Bpa	~10%
K153Bpa	~10%
K156Bpa	ND*
E162Bpa	~5%
E214Bpa	ND
K222Bpa	~5%
I308Bpa	30%
L309Bpa	30%
K310Bpa	30%
D312Bpa	30%

*ND - not detected due to low binding to mtRNAP

C

TEFM	A-T activity
WT	100%
sub140-144	50%
K142A/K144A	95%
sub149-153	5%
K153A/K156A	100%
F150A/L151G	100%
F150S/L151S	100%
R149A/R152A	80%
R152A/R153A	40%
K137Bpa	NT*
R140Bpa	NT
K142Bpa	NT
S145Bpa	100%
F150Bpa	NT
K153Bpa	NT
K156Bpa	<50%

* not tested

Figure 36. List of cross-linked peptides detected and activity of TEFM mutants.

(A) List of peptides representing inter-protein cross-links between mtRNAP and TEFM detected by BS3 (black), EDC (blue) and DSG (red). **(B)** List of pBpa-substitutions made in mtRNAP and TEFM with their respective cross-linking efficiencies. **(C)** Anti-termination activity of various TEFM mutants.

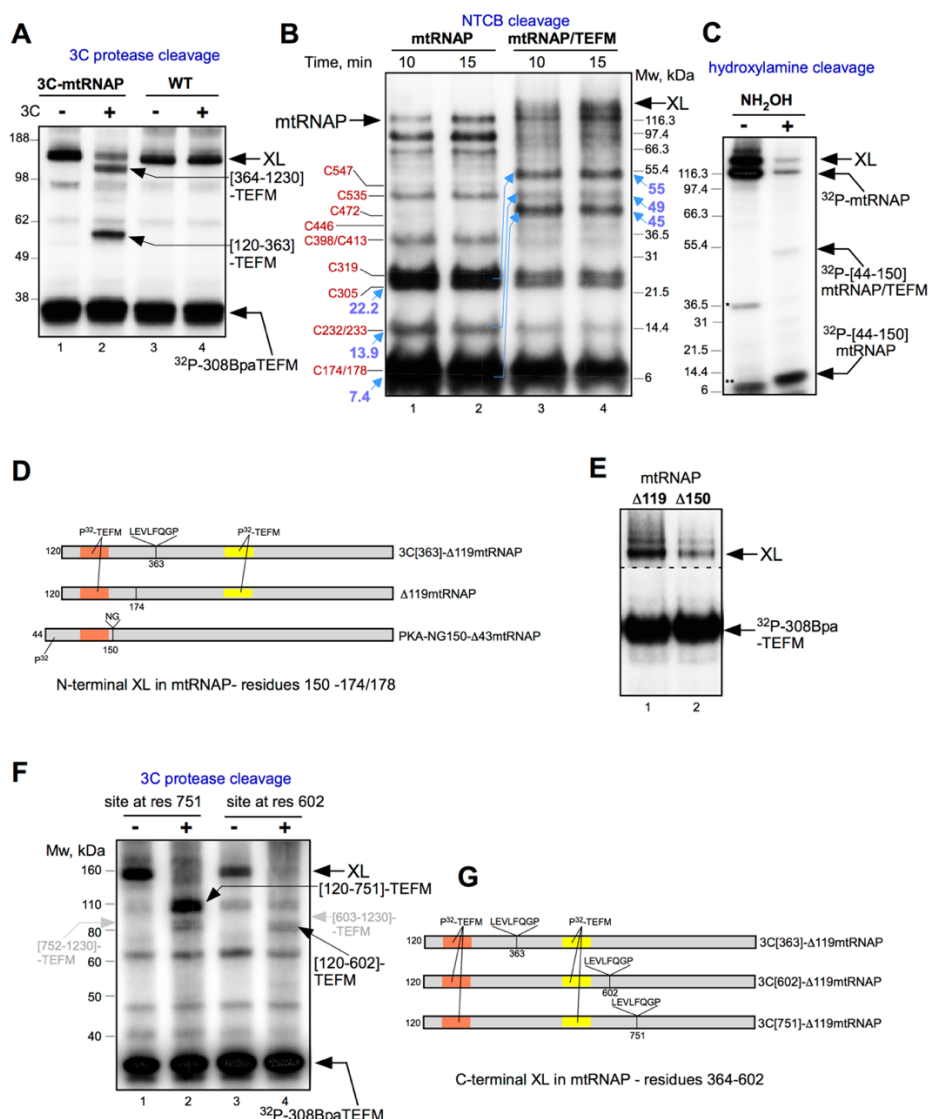


Figure 37. Mapping of interactions between TEFM and mtRNAP.

(A) The docking loop of TEFM cross-links to two distinct sites in mtRNAP. Cross-linking was performed using a ³²P-labeled Bpa308 TEFM variant and mtRNAP having an engineered cleavage site for the 3C protease introduced into the proline-rich linker domain (cleavage after residue 363). The cleavage reaction revealed that the majority of the cross-link to the 308BpaTEFM (~65%) is located to the N-terminus of the 3C site, i.e. in the N-terminal extension region/PPR domain of mtRNAP (residues 120-364). (B) Fine mapping of the N-terminal cross-linking site in mtRNAP. Cross-linking was performed using Bpa308 TEFM and ³²P-labeled Δ119 mtRNAP. Upon cross-linking, the radioactive species were separated using 6% SDS PAGE. MtRNAP-TEFM cross-link and mtRNAP were excised from the gel, eluted and subjected to cleavage by NTCB (cuts at Cys residues). The products of the reaction were resolved using 4-12% PAGE (Invitrogen). N-terminal cleavage pattern is observed for mtRNAP (lanes 1,2); the identity of every band representing labeled peptide was verified using the molecular weight standards (Mark 12, Invitrogen). Cleavage of the mtRNAP-TEFM cross-link results in appearance of adducts (indicated by the blue arrows), that are a TEFM mass (~37 kDa) larger than the corresponding peptides observed for mtRNAP. The cleavage pattern of the cross-link (lanes 3,4) is consistent with the location of the N-terminal cross-link between residues 120-178 in mtRNAP. Blue numbers (left side of the gel) represent apparent molecular weights of the three most N-terminal mtRNAP peptides involved in TEFM cross-linking; blue numbers to the right are molecular weights of the corresponding adducts. Note that the large molecular weight C-terminal cross-linked species are not resolved in this gel. (C) Fine mapping of the TEFM-mtRNAP NTD cross-link. The cross-linking was performed using a Bpa308 TEFM variant and ³²P-labeled Δ44 mtRNAP (mature mtRNAP) with the engineered hydroxylamine cleavage site (NG) at residue 150. Upon the cleavage reaction, appearance

of the labeled species (~55 kDa) corresponding to TEFM cross-linked to the region 44-150 in the NTD of mtRNAP was observed. Taken together with the data obtained in the NTCB experiment above, this suggests that TEFM interacts with the region 120-150 in the mtRNAP NTD. Note that the C-terminal cross-link is not visible in this experiment as the labeling of mtRNAP is due to the engineered PKA site at the N-terminus of mtRNAP. **(D)** Schematic illustration of the variants of mtRNAP used in experiments shown in Panel A-C. The red rectangle indicates the N-terminal cross-link to ^{32}P -labeled TEFM (or ^{32}P -labeled mtRNAP to TEFM); The yellow rectangle represents the C-terminal mtRNAP cross-link due to ^{32}P -labeled TEFM. **(E)** TEFM does not efficiently cross-link to the $\Delta 150$ mtRNAP. Cross-linking was performed using a ^{32}P -labeled Bpa308 TEFM variant and $\Delta 119$ (lane 1) or $\Delta 150$ (lane 2) mtRNAP. **(F)** Mapping of the C-terminal cross-linking site in mtRNAP. Cross-linking was performed using a ^{32}P -labeled Bpa308 TEFM variant and mtRNAPs with the engineered cleavage site for the 3C protease introduced into the thumb domain (cleavage after residue 751, lanes 1,2) or into the lever loop (cleavage after residue 602, lanes 3,4). The identified cross-linked adducts are indicated by black arrows. Grey arrows mark the expected position of the cross-linked species that correspond to the minor (< 5-10%) cross-linking sites. Cleavage of 3C-751-mtRNAP results in appearance of a single band, representing the region 120-751 of mtRNAP. Cleavage of 3C-602-mtRNAP results in appearance of a band, which represents the region 120-602 of mtRNAP. These data, taken together with the data suggesting the location of the C-terminal cross-linking site between residues 365-1230 (panel A), indicate the location of the C-terminal crosslink in region 364-602 of mtRNAP. **(G)** Schematic illustration of the variants of mtRNAP used in experiments shown in Panel A and F. The red rectangle indicates the N-terminal cross-link to ^{32}P -labeled TEFM; the yellow rectangle represents the C-terminal mtRNAP cross-link due to ^{32}P -labeled TEFM.

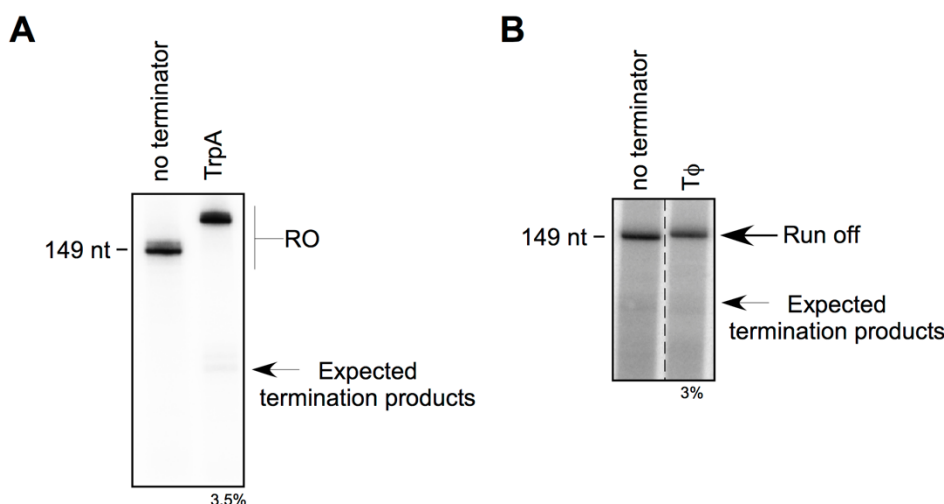


Figure 38. MtRNAP does not terminate at TrpA and TΦ hairpin terminators.

(A) Transcription assay performed using a PCR template containing the LSP promoter (lane 1) and the LSP promoter and TrpA terminator (lane 2). The position of the expected termination product is indicated with the black arrow; efficiency of termination at this point is indicated beneath the gel. **(B)** Transcription assay performed using a PCR template containing the LSP promoter (lane 1) and the LSP promoter and TΦ terminator (lane 2). The position of the expected termination product is indicated with the black arrow; efficiency of termination at this point is indicated beneath the gel.

Table 16. X-Ray data collection and refinement statistics for human TEFM.

	TEFM CTD native	TEFM CTD SeMet-SAD^b	TEFM NTD native	TEFM NTD S-SAD
Data collection				
Space group	<i>C2</i>	<i>C2</i>	<i>P4₃2₁2</i>	<i>P4₃2₁2</i>
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	98.48, 112.54, 88.84	97.57, 112.79, 88.84	47.67, 47.67, 93.40	47.72, 47.72, 93.32
α , β , γ (°)	90.00, 110.13, 90.00	90, 109.76, 90.00	90, 90, 90	90, 90, 90
Wavelength	1.0000	0.9765	1.0000	2.0664
Resolution (Å) ^a	46.23 – 1.90 (1.94 – 1.90) ^a	48.32 – 2.68 (2.74 – 2.68)	42.46 – 1.30 (1.33 – 1.30)	42.49 – 2.10 (2.14 – 2.10)
<i>R</i> _{meas}	0.056 (2.78)	0.24 (3.50)	0.107 (4.67)	0.096 (2.51)
<i>I</i> / σ (<i>I</i>)	17.92 (0.79)	9.45 (0.62)	20.26 (0.60)	41.67 (7.74)
<i>CC</i> _{1/2}	100.0 (35.3)	99.7 (32.6)	100.0 (41.3)	100.0 (95.1)
Completeness (%)	98.1 (89.5)	99.7 (98.8)	99.7 (98.1)	99.4 (96.1)
Redundancy	6.60 (5.04)	11.86 (6.25)	25.05 (24.01)	37.62 (25.64)
Refinement				
Resolution (Å)	46.23 – 1.90		42.46 – 1.30	
No. reflections	70117		27057	
<i>R</i> _{work} / <i>R</i> _{free}	0.195 / 0.228		0.190 / 0.197	
No. atoms				
Protein	6724		692	
Ligands	24		4	
Water	129		80	
<i>B</i> factors (Å ²)				
Protein	67.33		27.16	
Ligands	79.72		56.01	
Water	57.78		38.21	
r.m.s deviations				
Bond lengths (Å)	0.003		0.009	
Bond angles (°)	0.51		0.95	
Ramachandran				
Preferred/allowed/ disallowed (%)	97.5 / 2.5 / 0.0		98.7 / 1.3 / 0.0	

^a Values in parentheses are for highest-resolution shell.^b The data reflect the merged results from several crystals

Table 17. X-Ray data collection and refinement statistics for the EC-TEFM complex.

EC-TEFM Complex ^b	
Data collection	
Space group	C2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	224.53, 155.55, 164.19
α , β , γ (°)	90.00, 113.59, 90.00
Wavelength	1.0000
Resolution (Å) ^a	49.16 – 3.90 (4.00 – 3.90) ^a
<i>R</i> _{meas}	0.258 (2.187)
<i>I</i> / σ (<i>I</i>)	9.72 (1.44)
<i>CC</i> _{1/2}	99.7 (54.5)
Completeness (%)	99.6 (97.7)
Redundancy	13.12 (12.50)
Refinement	
Resolution (Å)	48.87 – 3.90
No. reflections	46930
<i>R</i> _{work} / <i>R</i> _{free}	0.243 / 0.276
No. atoms	
Macromolecules	25122
Ligands	0
Water	0
<i>B</i> factors (Å ²)	
Macromolecules	160.94
Ligands	0
Water	0
r.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.63
Ramachandran	
Preferred/allowed/ disallowed (%)	94.2 / 5.1 / 0.6

^a Values in parentheses are for highest-resolution shell.^b The data reflect the merged results from two datasets.

References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Agaronyan, K., Morozov, Y.I., Anikin, M., and Temiakov, D. (2015). Replication-transcription switch in human mitochondria. *Science* 347, 548–551.
- Alam, T.I., Kanki, T., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., Takio, K., Hamasaki, N., and Kang, D. (2003). Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Research* 31, 1640–1645.
- Aloni, Y., and Attardi, G. (1971a). Expression of the mitochondrial genome in HeLa cells. II. Evidence for complete transcription of mitochondrial DNA. *J. Mol. Biol.* 55, 251–267.
- Aloni, Y., and Attardi, G. (1971b). Symmetrical in vivo transcription of mitochondrial DNA in HeLa cells. *Proc. Natl. Acad. Sci. U.S.A.* 68, 1757–1761.
- Amunts, A., Brown, A., Toots, J., Scheres, S.H.W., and Ramakrishnan, V. (2015). Ribosome. The structure of the human mitochondrial ribosome. *Science* 348, 95–98.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465.
- Andersson, S.G.E., Karlberg, O., Canbäck, B., and Kurland, C.G. (2003). On the origin of mitochondria: a genomics perspective. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 165–77–discussion177–9.
- Ariyoshi, M., Vassilyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994). Atomic structure of the RuvC resolvase: A holliday junction-specific endonuclease from *E. coli*. *Cell* 78, 1063–1072.
- Arnold, J.J., Sharma, S.D., Feng, J.Y., Ray, A.S., Smidansky, E.D., Kireeva, M.L., Cho, A., Perry, J., Vela, J.E., Park, Y., et al. (2012). Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. *PLoS Pathog.* 8, e1003030.
- Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of Pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 79, 137–158.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209–1211.
- Baltimore, D., and Franklin, R.M. (1962). Preliminary data on a virus-specific enzyme system responsible for the synthesis of viral RNA. *Biochemical and Biophysical Research Communications* 9, 388–392.

- Barrell, B.G., Bankier, A.T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature* 282, 189–194.
- Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research* 41, W349–W357.
- Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716.
- Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Research* 44, D1251–D1257.
- Cermakian, N., Ikeda, T.M., Cedergren, R., and Gray, M.W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Research* 24, 648–654.
- Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W., and Cedergren, R. (1997). On the evolution of the single-subunit RNA polymerases. *J. Mol. Evol.* 45, 671–681.
- Ceschini, S. (2001). Crystal structure of the fission yeast mitochondrial Holliday junction resolvase Ydc2. *Embo J.* 20, 6601–6611.
- Chamberlin, M., McGrath, J., and Waskell, L. (1970). New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature* 228, 227–231.
- Chang, D.D., and Clayton, D.A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc. Natl. Acad. Sci. U.S.A.* 82, 351–355.
- Cheetham, G.M., and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286, 2305–2309.
- Cheetham, G.M., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80–83.
- Chen, X.J., and Butow, R.A. (2005). The organization and inheritance of the mitochondrial genome. *Nat. Rev. Genet.* 6, 815–825.
- Cheung, A.C.M., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. *Embo J.* 30, 4755–4763.
- Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the *S. cerevisiae* Spt6 core and C-terminal tandem SH2 domain. *J. Mol. Biol.* 408, 697–713.
- Cramer, P. (2002a). Common structural features of nucleic acid polymerases. *Bioessays* 24, 724–729.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–1876.
- Cramer, P., Bushnell, D.A., Fu, J., Gnat, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.
- Cramer, P. (2002b). Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* 12, 89–97.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563.

- Crick, F.H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* *12*, 138–163.
- Dairaghi, D.J., Shadel, G.S., and Clayton, D.A. (1995a). Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation. *Biochim. Biophys. Acta* *1271*, 127–134.
- Dairaghi, D.J., Shadel, G.S., and Clayton, D.A. (1995b). Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator. *J. Mol. Biol.* *249*, 11–28.
- Doe, C.L., Osman, F., Dixon, J., and Whitby, M.C. (2000). The Holliday junction resolvase SpCCE1 prevents mitochondrial DNA aggregation in *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* *263*, 889–897.
- Dong, A., Xu, X., Edwards, A.M., Midwest Center for Structural Genomics, Structural Genomics Consortium, Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., et al. (2007). In situ proteolysis for protein crystallization and structure determination. *Nat. Methods* *4*, 1019–1021.
- Doubl  , S. (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* *363*, 91–108.
- Durniak, K.J., Bailey, S., and Steitz, T.A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* *322*, 553–557.
- Ekstrand, M.I., Falkenberg, M., Rantanen, A., Park, C.B., Gaspari, M., Hultenby, K., Rustin, P., Gustafsson, C.M., and Larsson, N.-G. (2004). Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum. Mol. Genet.* *13*, 935–944.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 486–501.
- Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., Bernecky, C., and Cramer, P. (2017). Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* *169*, 120–131.e122.
- Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An allosteric path to transcription termination. *Molecular Cell* *28*, 991–1001.
- Ernster, L., and Schatz, G. (1981). Mitochondria: a historical review. *J. Cell Biol.* *91*, 227s–255s.
- Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.-G., and Gustafsson, C.M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat. Genet.* *31*, 289–294.
- Falkenberg, M., Larsson, N.-G., and Gustafsson, C.M. (2007). DNA replication and transcription in mammalian mitochondria. *Annu. Rev. Biochem.* *76*, 679–699.
- Fan, L., Sanschagrin, P.C., Kaguni, L.S., and Kuhn, L.A. (1999). The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: implications for a dual role as a primer recognition factor and processivity clamp. *Proc. Natl. Acad. Sci. U.S.A.* *96*, 9527–9532.
- Fan, L., Kim, S., Farr, C.L., Schaefer, K.T., Randolph, K.M., Tainer, J.A., and Kaguni, L.S. (2006). A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase. *J. Mol. Biol.* *358*, 1229–1243.
- Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* *147*, 1257–1269.

- Feng, J.Y., Xu, Y., Barauskas, O., Perry, J.K., Ahmadyar, S., Stepan, G., Yu, H., Babusis, D., Park, Y., McCutcheon, K., et al. (2016). Role of Mitochondrial RNA Polymerase in the Toxicity of Nucleotide Inhibitors of Hepatitis C Virus. *Antimicrob. Agents Chemother.* *60*, 806–817.
- Fisher, R.P., and Clayton, D.A. (1988). Purification and characterization of human mitochondrial transcription factor 1. *Mol. Cell. Biol.* *8*, 3496–3509.
- Fisher, R.P., Lisowsky, T., Parisi, M.A., and Clayton, D.A. (1992). DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein. *J. Biol. Chem.* *267*, 3358–3367.
- Fisher, R.P., Topper, J.N., and Clayton, D.A. (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* *50*, 247–258.
- Friedman, J.R., and Nunnari, J. (2014). Mitochondrial form and function. *Nature* *505*, 335–343.
- Fusté, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y., Atanassova, N., Holt, I.J., Gustafsson, C.M., and Falkenberg, M. (2010). Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Molecular Cell* *37*, 67–78.
- Gaspari, M., Falkenberg, M., Larsson, N.-G., and Gustafsson, C.M. (2004). The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *Embo J.* *23*, 4606–4614.
- Gillum, A.M., and Clayton, D.A. (1979). Mechanism of mitochondrial DNA replication in mouse L-cells: RNA priming during the initiation of heavy-strand synthesis. *J. Mol. Biol.* *135*, 353–368.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* *292*, 1876–1882.
- Górecka, K.M., Komorowska, W., and Nowotny, M. (2013). Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Research* *41*, 9945–9955.
- Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. *Science* *283*, 1476–1481.
- Greber, B.J., and Ban, N. (2016). Structure and Function of the Mitochondrial Ribosome. *Annu. Rev. Biochem.* *85*, 103–132.
- Greber, B.J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D., and Ban, N. (2015). Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* *348*, 303–308.
- Guja, K.E., Venkataraman, K., Yakubovskaya, E., Shi, H., Mejia, E., Hambardjieva, E., Karzai, A.W., and Garcia-Diaz, M. (2013). Structural basis for S-adenosylmethionine binding and methyltransferase activity by mitochondrial transcription factor B1. *Nucleic Acids Research* *41*, 7947–7959.
- Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Molecular Cell* *3*, 495–504.
- Gusarov, I., and Nudler, E. (2001). Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* *107*, 437–449.
- Gustafsson, C.M., Falkenberg, M., and Larsson, N.-G. (2016). Maintenance and Expression of Mammalian Mitochondrial DNA. *Annu. Rev. Biochem.* *85*, 133–160.
- Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* *12*, 483–492.

- Hamanaka, R.B., and Chandel, N.S. (2010). Mitochondrial reactive oxygen species regulate cellular signaling and dictate biological outcomes. *Trends in Biochemical Sciences* 35, 505–513.
- Hällberg, B.M., and Larsson, N.-G. (2014). Making Proteins in the Powerhouse. *Cell Metabolism* 20, 226–240.
- Heijne, von, G. (1986). Why mitochondria need a genome. *FEBS Lett.* 198, 1–4.
- Helmann, J.D., and Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57, 839–872.
- Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* 399, 17–30.
- Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36, 39–56.
- Hippel, von, P.H., Bear, D.G., Morgan, W.D., and McSwiggen, J.A. (1984). Protein-nucleic acid interactions in transcription: a molecular analysis. *Annu. Rev. Biochem.* 53, 389–446.
- Holt, I.J., Lorimer, H.E., and Jacobs, H.T. (2000). Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100, 515–524.
- Holt, I.J., and Reyes, A. (2012). Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol* 4, a012971–a012971.
- Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from *Pseudomonas aeruginosa*. *J. Mol. Biol.* 377, 1460–1473.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.
- Kang, D., Miyako, K., Kai, Y., Irie, T., and Takeshige, K. (1997). In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction. *J. Biol. Chem.* 272, 15275–15279.
- Kang, E., Wu, J., Gutierrez, N.M., Koski, A., Tippner-Hedges, R., Agaronyan, K., Platero-Luengo, A., Martinez-Redondo, P., Ma, H., Lee, Y., et al. (2016). Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature*.
- Kempken, F., Hermanns, J., and Osiewacz, H.D. (1992). Evolution of linear plasmids. *J. Mol. Evol.* 35, 502–513.
- Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular Cell* 16, 955–965.
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256, 1783–1790.
- Kroemer, G., and Reed, J.C. (2000). Mitochondrial control of cell death. *Nat. Med.* 6, 513–519.
- Kukat, C., Davies, K.M., Wurm, C.A., Spähr, H., Bonekamp, N.A., Kühl, I., Joos, F., Polosa, P.L., Park, C.B., Posse, V., et al. (2015). Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11288–11293.

- Kukat, C., Wurm, C.A., Spähr, H., Falkenberg, M., Larsson, N.-G., and Jakobs, S. (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 13534–13539.
- Kuzmine, I., Gottlieb, P.A., and Martin, C.T. (2001). Structure in nascent RNA leads to termination of slippage transcription by T7 RNA polymerase. *Nucleic Acids Research* *29*, 2601–2606.
- Kühlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* *343*, 1443–1444.
- Lane, W.J., and Darst, S.A. (2010a). Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* *395*, 671–685.
- Lane, W.J., and Darst, S.A. (2010b). Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.* *395*, 686–704.
- Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* *387*, 493–497.
- Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010). Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell Proteomics* *9*, 1634–1649.
- Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M., and Temiakov, D. (2010). Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. *J. Biol. Chem.* *285*, 18129–18133.
- Liu, B., Zuo, Y., and Steitz, T.A. (2015). Structural basis for transcription reactivation by RapA. *Proc. Natl. Acad. Sci. U.S.A.* *112*, 2006–2010.
- Lockshon, D., Zweifel, S.G., Freeman-Cook, L.L., Lorimer, H.E., Brewer, B.J., and Fangman, W.L. (1995). A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* *81*, 947–955.
- Ma, K., Temiakov, D., Anikin, M., and McAllister, W.T. (2005). Probing conformational changes in T7 RNA polymerase during initiation and termination by using engineered disulfide linkages. *Proc. Natl. Acad. Sci. U.S.A.* *102*, 17612–17617.
- Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* *12*, 783–793.
- Masters, B.S., Stohl, L.L., and Clayton, D.A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* *51*, 89–99.
- Matsunaga, M., and Jaehning, J.A. (2004). Intrinsic promoter recognition by a “core” RNA polymerase. *J. Biol. Chem.* *279*, 44239–44242.
- McAllister, W.T., and Raskin, C.A. (1993). The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Molecular Microbiology* *10*, 1–6.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* *40*, 658–674.
- McCulloch, V., Seidel-Rogol, B.L., and Shadel, G.S. (2002). A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol. Cell. Biol.* *22*, 1116–1125.

- Metodiev, M.D., Lesko, N., Park, C.B., Cámara, Y., Shi, Y., Wibom, R., Hultenby, K., Gustafsson, C.M., and Larsson, N.-G. (2009). Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome. *Cell Metabolism* 9, 386–397.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc. Natl. Acad. Sci. U.S.A.* 98, 892–897.
- Minczuk, M., He, J., Duch, A.M., Ettema, T.J., Chlebowski, A., Dzionek, K., Nijtmans, L.G.J., Huynen, M.A., and Holt, I.J. (2011). TEFM (c17orf42) is necessary for transcription of human mtDNA. *Nucleic Acids Research* 39, 4284–4299.
- Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and Attardi, G. (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 79, 7195–7199.
- Montoya, J., Gaines, G.L., and Attardi, G. (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* 34, 151–159.
- Mootha, V.K., Bunkenborg, J., Olsen, J.V., Hjerrild, M., Wisniewski, J.R., Stahl, E., Bolouri, M.S., Ray, H.N., Sihag, S., Kamal, M., et al. (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629–640.
- Morozov, Y.I., and Temiakov, D. (2016). Human Mitochondrial Transcription Initiation Complexes Have Similar Topology on the Light and Heavy Strand Promoters. *J. Biol. Chem.* 291, 13432–13435.
- Morozov, Y.I., Agaronyan, K., Cheung, A.C.M., Anikin, M., Cramer, P., and Temiakov, D. (2014). A novel intermediate in transcription initiation by human mitochondrial RNA polymerase. *Nucleic Acids Research*.
- Morozov, Y.I., Parshin, A.V., Agaronyan, K., Cheung, A.C.M., Anikin, M., Cramer, P., and Temiakov, D. (2015). A model for transcription initiation in human mitochondria. *Nucleic Acids Research* 43, 3726–3735.
- Mukundan, V.T., and Phan, A.T. (2013). Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* 135, 5017–5028.
- Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* 13, 31–39.
- Neupert, W. (1997). Protein import into mitochondria. *Annu. Rev. Biochem.* 66, 863–917.
- Ngo, H.B., Kaiser, J.T., and Chan, D.C. (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nature Publishing Group* 18, 1290–1296.
- Ngo, H.B., Lovely, G.A., Phillips, R., and Chan, D.C. (2014). Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat Comms* 5.
- Nunnari, J., and Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell* 148, 1145–1159.
- Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470–474.
- Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* 313, 762–766.

- Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* 37, 843–844.
- Paratkar, S., and Patel, S.S. (2010). Mitochondrial transcription factor Mtf1 traps the unwound non-template strand to facilitate open complex formation. *J. Biol. Chem.* 285, 3949–3956.
- Parisi, M.A., and Clayton, D.A. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* 252, 965–969.
- Perocchi, F., Gohil, V.M., Girgis, H.S., Bao, X.R., McCombs, J.E., Palmer, A.E., and Mootha, V.K. (2010). MICU1 encodes a mitochondrial EF hand protein required for Ca(2+) uptake. *Nature* 467, 291–296.
- Pham, X.H., Farge, G., Shi, Y., Gaspari, M., Gustafsson, C.M., and Falkenberg, M. (2006). Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J. Biol. Chem.* 281, 24647–24652.
- Phillips, A.F., Millet, A.R., Tigano, M., Dubois, S.M., Crimmins, H., Babin, L., Charpentier, M., Piganeau, M., Brunet, E., and Sfeir, A. (2017). Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion. *Molecular Cell* 65, 527–538.e6.
- Piskur, J. (1997). The transmission disadvantage of yeast mitochondrial intergenic mutants is eliminated in the mgt1 (cce1) background. *J. Bacteriol.* 179, 5614–5617.
- Pleiner, T., Bates, M., Trakhanov, S., Lee, C.-T., Schliep, J.E., Chug, H., Böhning, M., Stark, H., Urlaub, H., and Görlich, D. (2015). Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *Elife* 4, e11349.
- Posse, V., Shahzad, S., Falkenberg, M., Hallberg, B.M., and Gustafsson, C.M. (2015). TEFM is a potent stimulator of mitochondrial transcription elongation in vitro. *Nucleic Acids Research*.
- Posse, V., and Gustafsson, C.M. (2016). Human Mitochondrial Transcription Factor B2 is Required for Promoter Melting During Initiation of Transcription. *J. Biol. Chem.* jbc.M116.751008.
- Posse, V., Hoberg, E., Dierckx, A., Shahzad, S., Koolmeister, C., Larsson, N.-G., Wilhelmsson, L.M., Hällberg, B.M., and Gustafsson, C.M. (2014). The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation. *Nucleic Acids Research* 1–10.
- Pozzan, T., and Rizzuto, R. (2000). High tide of calcium in mitochondria. *Nat. Cell Biol.* 2, E25–E27.
- Ramachandran, A., Basu, U., Sultana, S., Nandakumar, D., and Patel, S.S. (2016). Human mitochondrial transcription factors TFAM and TFB2M work synergistically in promoter melting during transcription initiation. *Nucleic Acids Research*.
- Read, R.J., and McCoy, A.J. (2011). Using SAD data in Phaser. *Acta Crystallogr. D Biol. Crystallogr.* 67, 338–344.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.-K., Hagen, G., Guilfoyle, T.J., Pasa-Tolić, L., and Pikaard, C.S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Molecular Cell* 33, 192–203.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. *Nature* 478, 269–273.

- Rivera-Santiago, R.F., Sriswasdi, S., Harper, S.L., and Speicher, D.W. (2015). Probing structures of large protein complexes using zero-length cross-linking. *Methods* 89, 99–111.
- Robberson, D.L., Kasamatsu, H., and Vinograd, J. (1972). Replication of mitochondrial DNA. Circular replicative intermediates in mouse L cells. *Proc. Natl. Acad. Sci. U.S.A.* 69, 737–741.
- Roe, S.M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I.R., and Pearl, L.H. (1998). Crystal structure of an octameric RuvA-Holliday junction complex. *Molecular Cell* 2, 361–372.
- Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224, 234–237.
- Rubio-Cosials, A., Sidow, J.F., Jiménez-Menéndez, N., Fernández-Millán, P., Montoya, J., Jacobs, H.T., Coll, M., Bernadó, P., and Solà, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.* 18, 1281–1289.
- Saecker, R.M., Record, M.T., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* 412, 754–771.
- Schinkel, A.H., Koerkamp, M.J., Touw, E.P., and Tabak, H.F. (1987). Specificity factor of yeast mitochondrial RNA polymerase. Purification and interaction with core RNA polymerase. *J. Biol. Chem.* 262, 12785–12791.
- Schubot, F.D., Chen, C.J., Rose, J.P., Dailey, T.A., Dailey, H.A., and Wang, B.C. (2001). Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci.* 10, 1980–1988.
- Schwinghammer, K., Cheung, A.C.M., Morozov, Y.I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013). Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* 20, 1298–1303.
- Seidel-Rogol, B.L., McCulloch, V., and Shadel, G.S. (2003). Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nat. Genet.* 33, 23–24.
- Sheldrick, G.M., IUCr (2008). A short history of SHELX. *Acta Crystallogr Sect A Found Crystallogr* 64, 112–122.
- Shutt, T.E., and Gray, M.W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* 22, 90–95.
- Sigala, B., and Tsaneva, I.R. (2003). Functional dissection of the *Schizosaccharomyces pombe* Holliday junction resolvase Ydc2: in vivo role in mitochondrial DNA maintenance. *Eur. J. Biochem.* 270, 2837–2847.
- Sologub, M., Litonin, D., Anikin, M., Mustaev, A., and Temiakov, D. (2009). TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell* 139, 934–944.
- Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *Embo J.* 22, 2234–2244.
- Sousa, R. (1996). Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences* 21, 186–190.
- Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593–599.

- Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* 391, 231–232.
- Steitz, T.A., Smerdon, S.J., Jäger, J., and Joyce, C.M. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266, 2022–2025.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41, 207–234.
- Sugimoto, K., Kohara, Y., and Okazaki, T. (1987). Relative roles of T7 RNA polymerase and gene 4 primase for the initiation of T7 phage DNA replication in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 84, 3977–3981.
- Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (1999). Ubiquitin tag for sperm mitochondria. *Nature* 402, 371–372.
- Sweetser, D., Nonet, M., and Young, R.A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1192–1196.
- Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassilyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420, 43–50.
- Tan, B.G., Wellesley, F.C., Savery, N.J., and Szczelkun, M.D. (2016). Length heterogeneity at conserved sequence block 2 in human mitochondrial DNA acts as a rheostat for RNA polymerase POLRMT activity. *Nucleic Acids Research* gkw648.
- Tapper, D.P., and Clayton, D.A. (1981). Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. *J. Biol. Chem.* 256, 5109–5115.
- Taylor, S.W., Fahy, E., Zhang, B., Glenn, G.M., Warnock, D.E., Wiley, S., Murphy, A.N., Gaucher, S.P., Capaldi, R.A., Gibson, B.W., et al. (2003). Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.* 21, 281–286.
- Temiakov, D., Montesana, P.E., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W.T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14109–14114.
- Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassilyev, D.G. (2004). Structural basis for substrate selection by t7 RNA polymerase. *Cell* 116, 381–391.
- Tiranti, V., Savoia, A., Forti, F., D'Apolito, M.F., Centra, M., Rocchi, M., and Zeviani, M. (1997). Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. *Hum. Mol. Genet.* 6, 615–625.
- Uchida, A., Murugesapillai, D., Kastner, M., Wang, Y., Lodeiro, M.F., Prabhakar, S., Oliver, G.V., Arnold, J.J., Maher, L.J., Williams, M.C., et al. (2017). Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter. *Elife* 6, e27283.
- Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* 229, 105–124.
- Walberg, M.W., and Clayton, D.A. (1983). In vitro transcription of human mitochondrial DNA. Identification of specific light strand transcripts from the displacement loop region. *J. Biol. Chem.* 258, 1268–1275.

- Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M., and Gustafsson, C.M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Research* 40, 10334–10344.
- Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16072–16077.
- Wanrooij, S., Fusté, J.M., Farge, G., Shi, Y., Gustafsson, C.M., and Falkenberg, M. (2008). Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11122–11127.
- Washburn, R.S., and Gottesman, M.E. (2015). Regulation of transcription elongation and termination. *Biomolecules* 5, 1063–1078.
- Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98.
- Wilkins, M.H.F., Strokes, A.R., and Wilson, H.R. (2003). Molecular structure of deoxypentose nucleic acids. 1953.
- Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552.
- Wyatt, H.D.M., and West, S.C. (2014). Holliday junction resolvases. *Cold Spring Harb Perspect Biol* 6, a023192–a023192.
- Yakubovskaya, E., Mejia, E., Byrnes, J., Hambardjiev, E., and Garcia-Diaz, M. (2010). Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. *Cell* 141, 982–993.
- Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.-X., Chen, H.-F., et al. (2012). Identification of cross-linked peptides from complex samples. *Nat. Methods* 9, 904–906.
- Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H.T., and Holt, I.J. (2002). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* 111, 495–505.
- Yasukawa, T., Reyes, A., Cluett, T.J., Yang, M.Y., Bowmaker, M., Jacobs, H.T., and Holt, I.J. (2006). Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *Embo J.* 25, 5358–5371.
- Yin, Y.W., and Steitz, T.A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* 298, 1387–1395.
- Yin, Y.W., and Steitz, T.A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* 116, 393–404.
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811–824.
- Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural Basis of Transcription Initiation. *Science* 338, 1076–1080.

List of Abbreviations

Å	Angstrom
°C	degrees Celsius
ATP	adenosine triphosphate
ATP	adenosine-5'-triphosphate
AU	arbitrary units
BIS-TRIS	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
bp	base pair(s)
Bpa	4-benzoylphenylalanine
BS3	(bis(sulfosuccinimidyl)suberate)
CD	circular dichroism
CSB	Conserved Sequence Block
CTD	C-terminal domain
CTP	cytidine-5'-triphosphate
CV	column volume
Da	Dalton
ddH ₂ O	double-distilled H ₂ O
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSG	disuccinimidyl glutarate
DTT	dithiothreitol
<i>E.coli</i>	Escherichia coli
EC	elongation complex
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
<i>et al.</i>	and other [latin <i>et alii</i>]
EtOH	ethanol
g	gram or gravitational acceleration
GTP	guanosine-5'-triphosphate
h	hour(s)
<i>H.sapiens</i>	Homo sapiens
HCl	hydrochloric acid
His	histidine
HMG	high mobility group
HSP	heavy strand promoter
IC	initiation complex
kb	kilo base pair
kDa	kilo Dalton
LB	Luria Broth
LDS	lithium dodecyl sulfate
LSP	light strand promoter

M	molar (mol/liter)
MES	2-(n-morpholino)ethanesulfonic acid
Mg	magnesium
min	minutes(s)
ml	milliliter
MOPS	3-(n-morpholino)propanesulfonic acid
mRNA	messenger RNA
mtDNA	mitochondrial deoxyribonucleic acid
mtRNAP	mitochondrial DNA-dependent RNA polymerase
MTS	mitochondria transport signal
ND6	NADH dehydrogenase subunit 6
Ni-NTA	nickel-nitrilotriacetic acid
nt	nucleotide
NT	non-template strand
NTD	N-terminal domain
NTP	nucleoside triphosphate
OD ₆₀₀	optical density at 600 nm
O _H	heavy strand origin of replication
O _L	light strand origin of replication
<i>P.aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PAGE	polyacrylamide gel electrophoresis
PBD	promoter binding domain
PCR	polymerase chain reaction
PDB	Protein Data Base
PEG	poly(ethylene glycol)
PI	protease inhibitor mix
PMSF	phenylmethanesulfonyl fluoride
Pol I	DNA-dependent RNA polymerase I
Pol II	DNA-dependent RNA polymerase II
Pol III	DNA-dependent RNA polymerase III
Pol IV	DNA-dependent RNA polymerase IV
Pol V	DNA--dependent RNA polymerase V
PPi	pyrophosphate
PPR	pentatricopeptide repeat
preIC	pre-initiation complex
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAP	DNA-dependent RNA polymerase
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
s	second(s)
<i>S.cerevisiae</i>	<i>Saccharomyces cerevisiae</i>

<i>S.pombe</i>	<i>Saccharomyces pombe</i>
SAD	single-wavelength anomalous diffraction
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	selenomethionine
siRNA	small interfering RNA
ssRNAP	single-subunit DNA-dependent RNA polymerase
T7 RNAP	bacteriophage T7 DNA-dependent RNA polymerase
TAE	Tris-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TS	template strand
TSS	transcription start site
UTP	uridine-5'-triphosphate
V	Volt
v / v	volume per volume
w / v	weight per volume
WT	wild-type
XL	cross-link

List of Figures

Figure 1. The central dogma of molecular biology.....	1
Figure 2. Comparison of single and multi-subunit RNA polymerases.....	4
Figure 3. Organization of the human mitochondrial genome.	5
Figure 4. The sequential model of transcription initiation in human mitochondria.	8
Figure 5. Structures of human mitochondrial RNA polymerase.	10
Figure 6. An incomplete structural picture of human mitochondrial transcription.	12
Figure 7: Biochemical characterization and crystallization of a human TFB2M variant.....	38
Figure 8. Structure of human TFB2M.	39
Figure 9. Structural comparison of TFB2M homologs.....	40
Figure 10. Reconstitution and crystallization of the human mitochondrial IC.....	41
Figure 11. Structure determination of the human mitochondrial IC.....	42
Figure 12. Structure of the human mitochondrial transcription initiation complex.	44
Figure 13. Promoter DNA interactions in the IC and DNA bending.....	45
Figure 14. TFAM recruits mtRNAP to promoter DNA.....	46
Figure 15. Identification of the tether helix and architecture of the IC at HSP.	47
Figure 16. TFB2M stabilizes open DNA and traps the non-template strand.....	48
Figure 17. Structural basis of DNA melting and open DNA stabilization.	49
Figure 18. Comparison of human mitochondrial IC and T7 RNAP IC.....	51
Figure 19. Transition from the initiation phase to the elongation phase of transcription.....	52
Figure 20. Model for transcription initiation in human mitochondria.	54
Figure 21. Domain organization, crystallization and structure determination of human TEFM.	57
Figure 22. Crystal structure of human TEFM.....	59
Figure 23. Secondary structure topology of human TEFM.	60
Figure 24. Functional role of TEFM domains.	61
Figure 25. Topology of the anti-termination complex.....	63
Figure 26. Structure of the anti-termination complex.....	64
Figure 27. Dimerization of TEFM is required for EC binding and anti-termination activity.....	65
Figure 28. Properties of the anti-termination complex.	66
Figure 29. MtRNAP adopts the post-translocated conformation in the anti-termination complex.....	68
Figure 30. TEFM interactions revealed by cross-linking.	70
Figure 31. TEFM does not bind to the initiation complex.....	71
Figure 32. Mechanism of transcription termination at CSBII.	73
Figure 33. TEFM is a pseudo-resolvase.	77
Figure 34. Structural view on the human mitochondrial transcription cycle.....	78
Figure 35. TEFM interacts with RNA and DNA in the EC.....	92
Figure 36. List of cross-linked peptides detected and activity of TEFM mutants.	93
Figure 37. Mapping of interactions between TEFM and mtRNAP.....	94
Figure 38. MtRNAP does not terminate at TrpA and T Φ hairpin terminators.	95

List of Tables

Table 1. Bacterial strains used in this study.....	14
Table 2. Growth media for <i>E.coli</i>	14
Table 3. Additives to growth media.....	15
Table 4. Plasmids used in this study.....	15
Table 5. Primers used in this study.....	17
Table 6. Synthetic oligonucleotides used for reconstitution of transcription complexes.	19
Table 7. General solutions	20
Table 8. Buffers used for purification of mtRNAP.....	21
Table 9. Buffers used for purification of TFAM	21
Table 10. Buffers used for purification of TFB2M.....	21
Table 11. Buffers used for purification of TEFM.....	22
Table 12. X-Ray data collection and refinement statistics for TFB2M ^{cryst} structure.....	85
Table 13. X-Ray data collection, phasing and refinement statistics for IC LSP structure	86
Table 14. X-Ray data collection statistics for IC LSP Bromine derivatives.	87
Table 15. X-Ray data collection, phasing and refinement statistics for IC HSP structure.	88
Table 16. X-Ray data collection and refinement statistics for human TEFM.	96
Table 17. X-Ray data collection and refinement statistics for the EC-TEFM complex.	97